DOCUMENT RESUME

ED 065 140 LI 003 761

AUTHOR	Vilentchuk, Lydia, Ed.; Haimovic, Gila, Ed.
TITLE	ISLIC International Conference on Information Science: Proceedings (Tel Aviv, 29 August - 3 September, 1971). Volume 2.
INSTITUTION	Israel Society of Special Libraries and Information Centres, Tel Aviv.
SPONS AGENCY	National Center of Scientific and Technological Information, Tel Aviv (Israel).
PUB DATE	72
NOTE	432p.;(144 References)
EDRS PRICE	MF-$0.65 HC-$16.45
DESCRIPTORS	*Automation; Conference Reports; Evaluation; Information Retrieval; *Information Science; *Information Scientists; Information Services; *Relevance (Information Retrieval); *Reprography; Training

ABSTRACT

Information science is a young and rapidly growing field, embracing a wide range of subjects and activities. The conference, in ten technical sessions, attempted to cover various aspects of information work. The areas covered were: information analysis and information analysis centers; retrieval of information; selection, education and training of personnel; and, publishing and reprography. The proceedings volumes are arranged in order of sessions. Summaries of discussions follow the papers presented at each session. Volume two includes papers delivered under the following broad topic categories: evaluation of retrieval effectiveness; selection, education and training of personnel; publishing and reprography; commercially available services; and processing for automation. A summary of an open meeting of the International Federation for Documentation Study Committee, "Information for Industry," is included. The list of participants and name and author affiliation indexes for both volumes appear in this volume. (Volume one is LI 003 760.) (Author/SJ)

ISLIC

ISLIC International Conference on Information Science

TEL AVIV, 29 AUGUST – 3 SEPTEMBER, 1971

proceedings

VOLUME 2

edited by LYDIA VILENTCHUK
assistant editor Gila Haimovic

NATIONAL CENTER OF SCIENTIFIC AND TECHNOLOGICAL INFORMATION

Library of Congress Catalog Card Number: 72-000007
UDC Number: 002:061.3(100)

Published by the:
National Center of Scientific and Technological Information
84 Hachashmonaim Street, Tel Aviv

Printed in Israel
by Multigraph Ltd

table of contents

Volume Two

* not delivered orally

SESSION FIVE

evaluation of retrieval effectiveness

Authors of Papers

ON THE EFFICIENCY OF DIFFERENT SEARCH STRATEGIES

C. Köhler and C. Wagner
German Cancer Research Center, Heidelberg
Institute for Documentation, Information and Statistics

SYNOPSIS

Five well known search strategies in EDP have been compared with a newly developed technique. The latter proved more efficient, both with regard to the mean number of access points and the probability of a hit within the first five attempts.

For the task of information retrieval in computer-based systems different search strategies have been developed, the efficiency of which is to be compared (cf. Table 1):

V Evaluation of Retrieval Effectiveness

Search Strategy	Mean No. of Access Points	Efficiency of Search	Efforts for Updating
1.) Sequential Search	$\frac{N+1}{2}$	poor	very small
2.) Binary Search	$\frac{N+1}{N} \log_2 (N+1) - 1$	good	great
3.) Block Search	$\frac{N+n^2}{2n}$	mean	mean
4a.) Single Chaining	$\frac{N+1}{2}$	poor	small
4b.) Double Chaining	$1.4 \log_2 N$	good	small
5.) Distributed Key Search	$1.75 \log_2 N$	good	small
6.) Serial Jump Strategy	$\sqrt{\frac{N}{K}} - \frac{1}{2}$	good	mean

Table 1: Properties of some Search Strategies

1.) <u>Sequential Search</u>

In searching sequentially, a file of N elements is checked with a key-word record per record until the element searched for has been found. Previous sorting of the material is not required. The minimum number of comparisons equals 1, the maximum $v_{max} = N$, therefore the mean number $v_m = \frac{N+1}{2}$.

If M elements are searched for simultaneously, the mean number of attempts per element will be $v_m = \frac{N}{M}$.

2.) <u>Binary Search</u>

The precondition for using the binary search method is strict order and a knowledge of the number of elements in the file. At every step the part of the file which is to be searched is subdivided into 3 parts, namely

1. the element addressed for the purpose of comparison
2. any "smaller" elements
3. any "larger" elements.

The element addressed is compared with the concept to be searched for ("equal", "smaller", "larger"). According to the result of this comparison, the search is either completed (element has been found) or it is to be continued within the "smaller" or "larger" elements only. The search strategy is optimal if the remaining quantity of elements is halved at each step. In this case, the ranges of the jump form a monotonously decreasing sequence with the factor 1/2.

For the N elements of a file $N = 2^t - 1$; this leads to the mean number of access points

$$v_m = \left[\frac{N+1}{N} \cdot \log_2 (N+1)\right] - 1$$

3.) <u>Block Search</u>

Using the block search requires the subdivision of a sorted file into blocks of length D. The first comparison takes place with the last element of the first block. If the key-word to be searched for is "larger", the programme jumps to the last element of the second block where renewed comparison takes place. The procedure is continued until the addressed element is "equal" to or "smaller" than the key-word. If it is "equal", the desired element is found; otherwise, the addressed block is looked up sequentially until the desired element has been found.

In a file of N elements with block length = D there are N/D blocks. In order to address all these sequentially, a maximum of N/D steps is necessary. Within a block a maximum of D-1 steps is required to find a certain element. This leads to a mean number of steps

$$Y_m = \frac{N + D^2}{2D}$$

4.) Chaining Methods

There are two different chaining methods, the single chain and the double chain. In the single chain method the address of the subsequent element is stored in addition to each element. The entire file is searched element by element, using the chained address, until the element searched for has been found. The single chain method corresponds to the sequential search in principle; the mean number of search steps required is the same, namely

$$Y_m = \frac{N + 1}{2}$$

The double chain method is only possible in strictly hierarchical structures. In this method 2 addresses are stored simultaneously with each element - namely the one of a neighbouring element on the same level and the one of a neighbouring element on the next lower level. The mean length of search is

$$Y_m = 1.4 \log_2 N$$

5.) Distributed Key-Search

The distributed key search requires a strictly hierarchical order of the elements in the file. Each level is characterized by a number; each element of the file is unequivocally determined by the combination of all numbers. On the basis of the complete combination of the numbers any desired element can be retrieved.

With a symmetrical order of the material and a mean number of branches of about 4 there is an optimal mean number of access points of

$$Y_m = 1.24 \log_2 N$$

However, this optimal value is rarely reached, since the data collections are hardly ever optimally structured.

(476) BEIN
JAMBE
LEG

•
•
•
•
•

(512) BLASTOGENESE
TRANSFORMATION BLASTIQUE
BLASTOGENESIS

BLEI
PLOMB
LEAD

BLUT
SANG
BLOOD

BLUTDRUCK
PRESSION SANGUINE
BLOOD PRESSURE

(516) BLUTGEFÄSSE
VAISSEAUX SANGUINS
BLOOD VESSELS

BLUTGRUPPE
GROUPE SANGUIN
BLOOD GROUP

BLUTSERUM
SERUM SANGUIN
BLOOD SERUM

BLUTVOLUMEN
VOLUME SANGUIN
BLOOD VOLUME

•
•
•
•

(524) BOR
BORE
BORON

Fig. 1: Part of the Threelingual Thesaurus

V Evaluation of Retrieval Effectiveness

None of the above named search strategies appeared to us to be optimally suited to the practical task of translating key-words in a trilingual thesaurus (Fig.1) in alphabetical order into another language. We, therefore, developed a search algorithm of our own - the so-called serial jump-search. The application of this method presupposes setting up a correspondence table containing the addresses within groups of the first 2 letters (cf. Table 2).

First 2 Letters	Address
AA	1
AC	125
AN	258
BA	360
BE	452
CA	596
DA	727
.	.
.	.
.	.

Table 2: Correspondence Table for the Described Example

This method is to be briefly explained by the following example: The first step towards the solution of the task consists in comparing the 2 first letters of the word to be translated with the correspondence table; thereby the beginning address for the search is established.

The word "Blutgefässe" (blood vessels) is to be translated from German into French. The first two letters of the word "Blutgefässe" (BL) lie between BE and CA. According to the correspondence table, the search interval ranges from address 452 to address 595 (Fig.2).

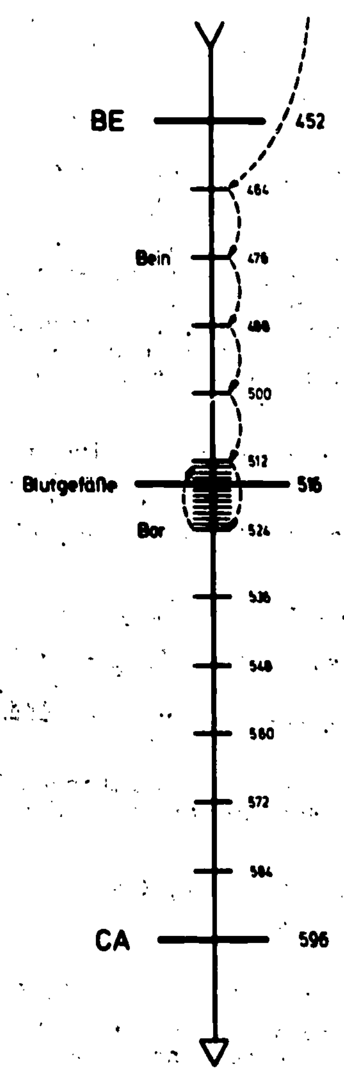


Fig. 2: Description of the Serial Jump Strategy

The number of elements to be compared from the start to the end of the search interval is called D. In our example D equals 144. D is subdivided into X segments (in our example X = 12). The first element of each segment (beginning with the second segment) is compared as to whether the word to be searched for is equal to, larger than or smaller than the one addressed. If the word to be searched for is equal, the search is completed. If it is larger than the word found, then the programme jumps to the first address of the next segment (in our case = 476) and this address is compared. The word "Bein" (leg) found at this address is smaller than the search word "Blutgefässe"; therefore, the search is continued at address 488 and so on. If the word to be searched for is smaller than the word found (e.g. "Blutgefässe" is smaller than "Bor" [Boron]), then the search is continued serially from the second word of the preceding segment until the word has been found within this segment.

The number of access points needed to find a certain word is composed of the number of jumps and the number of steps within a segment. (In our example 10 search steps were required, namely 6 segment jumps and 4 steps in the last segment.)
The general rule is:

minimum number of access points $Y_{min} = 1$

maximum number of access points $Y_{max} = (x-1)+(\frac{D}{x}-1) = x+\frac{D}{x}-2$

mean number of access points $Y_m = \frac{Y_{max}+Y_{min}}{2} = \frac{D}{2x} + \frac{x-1}{2}$ (1)

Dependent on the size of the search interval D, the value of X can be optimized by partial differentiation of the equation (1) ($X = \sqrt{D}$).

When inserting the values D = 144, X = 12 in equation (1) we obtain $Y_m = \frac{144}{24} + \frac{11}{2} = 11.5$.

In using D = 81 (X = 9) it follows that $Y_m = 8.5$.
With a total of 5.000 words in the thesaurus this value is already less than the corresponding value of the binary search.

From equation (1) it follows under optimal conditions that

$$Y_m = \frac{D}{2\sqrt{D}} + \frac{\sqrt{D}-1}{2} = \sqrt{D} - 1/2 \qquad (2)$$

With N = total number of elements in the file and S = number of intervals defined by the correspondence table

$$D = \frac{N}{S} \quad (3)$$

From (2) and (3) it follows that $Y_m = \sqrt{\frac{N}{S}} - 1/2$

This relation has been calculated for different values of N and S and is shown by a family of curves (Fig.3). Mean numbers of access points for the binary search are included for the purpose of comparison. Fig. 3 shows that our search strategy is better for a given value of N, if the number S of the address fields in the correspondence table is large enough.

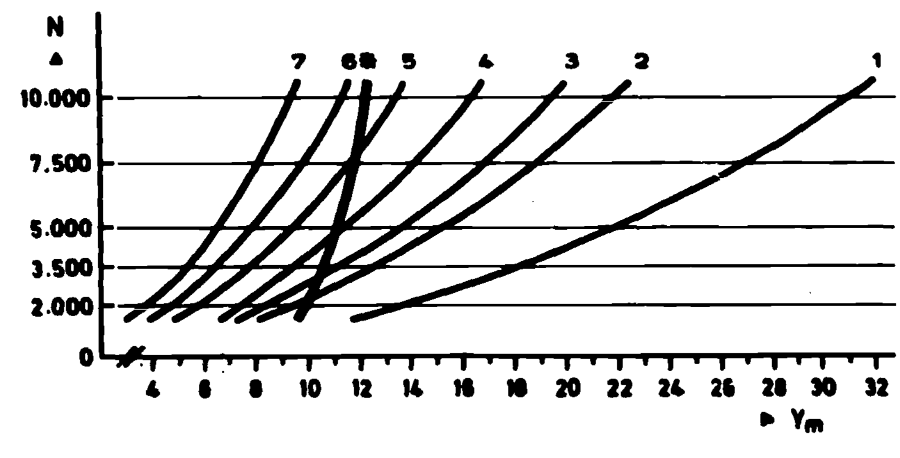


Fig. 3: The mean number of access points in serial jump strategy depending upon the size of the material to be searched and the number of address fields in the correspondence table compared with the corresponding values for the binary search.

Curve #	:	Binary Search
1	:	No. of address fields = 10
2	:	No. of address fields = 20
3	:	No. of address fields = 25
4	:	No. of address fields = 35
5	:	No. of address fields = 50
6	:	No. of address fields = 70
7	:	No. of address fields = 100

N = number of the elements of a file
Y_m = mean number of access points

EVALUATION OF RETRIEVAL EFFECTIVENESS

T.M. Aitchison and Angela M. Hall
INSPEC, Institution of Electrical Engineers, London

SYNOPSIS

A comparative evaluation is described of a number of controlled and 'natural' languages which were candidates for use for retrieval from the INSPEC data base and for SDI. The problems involved in such an evaluation and the general conclusions reached are discussed.

The comparative evaluation of index languages arose from the need to decide on the main subject organising and retrieval tool to be used in INSPEC's integrated computer-based system, one aspect of which is described by Simmons, Vernon and Martin in a paper in Area 4. Since the basis of the system is that each item included in the data base - at least 125,000 in 1971 - would be input only once, it was important to try and ensure that the index language used should be both the most suitable one available and equally useful for the printed indexes to the abstracts journals produced by INSPEC, for a selective dissemination of information (SDI) system, and for retrieval from the data base.

It is probably true to say that the choice was not completely open since a classification system (whether U.D.C., a true hierarchical one or a faceted classification) was never seriously considered and there was an implied assumption that whatever language was chosen should be employed in a coordinate-indexing mode.

Like most information workers we had had a fundamental conviction that a controlled language must be superior to an uncontrolled or 'natural' language, and that the performance of an information retrieval system was mainly dependent on the degree of intellectual effort involved in the setting up of the index language and in the indexing. However, the results of the second Cranfield tests (1) suggested that these convictions or innate assumptions might not be entirely valid, and that, at the very least, the whole question should be considered open.

It was on that basis that the possibility of conducting a comparative evaluation first presented itself. Although it was desired to select a language equally suitable for printed indexes, SDI and retrospective searching of a machine file, only for retrospective searching had a methodology been developed (by Cleverdon, Mills and Keen for the Cranfield (1) tests). This, although it had its share of criticism, seemed sensible, understandable and had the decided advantage that its critics had not been able to suggest an alternative. It was hoped to avoid at least some of the difficulties and forestall some of the criticism in the design of the new evaluation. In particular it was intended to take advantage of the INSPEC facilities by having software developed to allow the evaluation data to be obtained by program and by using, for the evaluation, documents already indexed for the normal production of INSPEC services.

The index or retrieval languages which it was decided should be studied, since they were already available or potentially available in the INSPEC database, comprised:

1. Title

2. Abstract and title

3. Hybrid: the subject heading and plain-language modifier line used in the printed indexes to the abstracts journal (i.e. a combination of controlled and uncontrolled language).

4. Free-index: free-language terms applied by an indexer (applied initially, in the SDI Investigation before translation into the controlled language).

5. Controlled: a thesaurus-based controlled language used in the INSPEC SDI Investigation.

It is obvious that the main problem in comparing these five options lies in ensuring that the controlled language is compared fairly with the other four languages in their free-language formulations. We hoped to overcome the problem by using a wide variety of free-language formulations and as many controlled-language formulations as possible.

Methodology

The design of the evaluation (2) was based on matching a test collection of documents with a test collection of questions to which each of the documents had been assessed for relevance. The documents in the collection required to be indexed by each of the languages to be evaluated, and the questions formulated in the same language.

The main difficulties that may be expected in setting up such an arrangement are that the document indexing for the test may be consciously or unconsciously biassed in favour of one of the languages; the questions or their formulation may favour one language; or it may be difficult to persuade the questioners to assess a large enough test collection of documents for relevance to their questions.

In the INSPEC evaluation it is considered that all of these difficulties were overcome, some more successfully than others.

In the indexing of documents the possibility of bias was entirely ruled out by selecting documents which had been routinely indexed before the test was postulated. The indexing had been carried out by different groups (i.e. Hybrid by one group and the two SDI languages by another) but it was considered that this was not a disadvantage since for INSPEC the main interest was in comparing the total system, i.e. both the language and its use.

The possibility of the choice of questions influencing the comparison was accepted rather than avoided by using 'real' questions submitted by a wide variety of electronics research workers, while the formulation difficulty was reduced, if not overcome, by using a total of 30 different formulations for each question, i.e. 10 question formulations, each in three different logical product coordinations or strategies.

The different question formulations used are shown in Figure 1. The three strategies required the coordination (by 'AND') of the minimum, a moderate, or the maximum number of terms in the question and corresponded to low, medium and high exhaustivity of search.

An example of the A and C formulations for a question is given in Figure 2.

In the assessment of documents for relevance to the questions, a method was found by which the questioner, with comparatively little effort was able to assess some 1200 documents. This depended on his having already assessed the documents for relevance to his profile in an SDI service and his being supplied with the profile-relevant documents for assessment of relevance to his question. The method worked satisfactorily in the main, although some assessments had to be discarded when it was found that some questioners had again assessed the question documents for relevance to their profile. A more serious consideration arises from the fact that the assessments were highly subjective, peculiar to each questioner, instead of objective and reproducible as would have been the case if a consensus judgement had been obtained. However it was considered that the subjective assessment represented the normal situation and should therefore be retained.

Question formulations

Single-term free-language formulations

A. The words of the question as stated by the questioner (and, where necessary, selected for the evaluation).

B. The words of 'A' with true synonyms and other word forms added as alternatives, and word endings confounded as appropriate.

C. As 'B' with the addition of quasi-synonyms and generic, more specific, or related terms as were considered likely to be advantageous.

Simple-concept free-language formulations

In these formulations, simple adjective-and-noun phrases, and a small number of others were preserved.

E. As 'A' with simple concepts preserved.

F. As 'B' " " " "

G. As 'C' " " " "

Hybrid (free-and controlled-language) formulations

D. As 'C' with appropriate Science Abstracts subject headings added as an additional set of search terms.

H. As 'D' but based on 'G'.

Controlled-language formulations

R. The words in the question translated by strict use of the thesaurus into the controlled language used in the SDI Investigation.

S. As 'R' but with free use of alternative terms, similar to the free-language formulations 'C' and 'G'.

Figure 1

An example of the A and C question formulations

Question: Information on the crystalline structure of cadmium sulphide?

Field	Formulation A	Formulation C
10	crystalline	cryst*
		xtal*
		microstructure*
11	structure	structur*
		orient*
		lattice*
		microstructur*
		state
12	cadmium	cadmium
		Cd
		Cds
13	sulphide	sulphide*
		Cds

Figure 2

DEVIL

To carry out the evaluation the DEVIL (Direct Evaluation of Index Languages) system was developed by INSPEC. By this system, which is based on the INSPEC general-purpose file maintenance system, the question and document data were input, validated and corrected: searches for matches were carried out as required: and (since the relevance assessments were included in the test collection files) the results were reported and the performance measures calculated by program for each formulation of each question and for the equivalent groups.

Matching tests

In the tests a total of 97 questions were used in the main set, and 82 in a subset from which questions with particularly doubtful (probably erroneous) relevance assessments had been removed. The document collection comprised 542 documents each 'indexed' by the five languages. An example of the indexing of an input document is given in Figure 3.

The evaluation consisted essentially of matching each formulation of each question against the appropriately-indexed document collection in two modes, for which the terms, Plain Coordinate Matching (PCM) and Boolean matching (Boolean) were coined. In the P.C.M. mode, a match on all the 'n' terms in the question was first required, then of any n-1, then of any n-2, etc. until single-term matches were reached. Such an output is shown in Figure 4.

In the alternative mode, Boolean, a logical relationship was specified for each question. Thus each question might require a different number of terms to be coordinated. An example of such an output is shown in Figure 5.

Three performance measures were calculated:

Recall (relating the number of relevant documents retrieved to the total number of relevant in the collection)

Precision (relating the number of relevant documents retrieved to the total number of documents retrieved)

Fallout (relating the number of non-relevant retrieved to the total non-relevant in the collection)

For each formulation of the set or subset of questions, the results for the individual questions were averaged by two methods and totalled by four different systems, since each of these had certain advantages but none was free from some drawback.

<u>**An example of the indexing of an input document**</u>

Title:	The ultrasonic interface micrometer
Abstract:	In nuclear reactors, thickness of the walls of tubes are measured by ultrasonic continuous-wave and pulse-wave micrometers. The principles and recording techniques of the instruments are described. The continuous-wave micrometer is used for simultaneous determination of thickness during defect-testing operations; the pulsed-wave instrument is preferred when better surface discrimination is needed.
Hybrid:	£BMEMD: Measurement by electrical methods, distance, nuclear reactor thickness, U.S. interferometry.
	£BMA: Measuring apparatus
	£BNP: Nuclear power
	£BU: Ultrasonics, tube thickness meas in nuclear reactor
Free-index:	Ultrasonic interference micrometer, thickness measurement DM, continuous wave micrometer, pulsed wave micrometer, design, recording, applications.
Controlled:	Ultrasonic waves, interference, dimensions, measurement, continuous waves, pulses, design, recording, applications.

Figure 3

V Evaluation of Retrieval Effectiveness

Plain coordinate matching output

Question: C030 Total documents: 542

Document indexing: Free-index Question descriptors: 4

Relevant documents: 2039/1, 2273/1, 3067/1, 3221/1

Coord-ination	Documents			Tot Ret	Rel Ret	R%	F%	P%
4	2039/1			1	1	25.0	0.0	100.0
3	3067/1	4137/0		3	2	50.0	0.2	66.7
2	1155/0	2273/1		5	3	75.0	0.4	60.0
1	1050/0	1081/1	1084/0					
	1097/0	1131/0	1133/0					
	1187/0	2005/0	2021/0					
	2030/0	2031/0	2048/0					
	2060/0	2166/0	2172/0					
	2219/0	2232/0	2235/0					
	2274/0	3082/0	3104/0					
	3214/0	3221/1	4117/0					
	4250/0	4287/0	4299/0	32	4	100.0	5.2	12.5

Figure 4

Boolean Output

Question formulation: C Total documents: 542

Document indexing: Free-index

Question	Documents Retrieved	Tot Ret	Tot Rel	Rel Ret	R%	F%	P%
C083	3116/0, 3246/1	2	1	1	50.0	0.2	50.0
C084		0	3	0	0.0	0.0	*
C085	3003/1	1	3	1	33.3	0.0	100.0
C086	2011/1, 2210/0	2	1	1	100.0	0.2	50.0

Figure 5

Results

The matching of the 542 documents and 97 questions in 10 formulations by 3 strategies, and the use of two methods of averaging and four types of totalling produced a very large quantity of data (3). Examples of Recall/Precision plots of P.C.M. and Boolean outputs are shown in Figures 6 and 7 respectively.

The data were supplemented by analyses of failures in recall and precision (i.e. why relevant documents were not retrieved and why non-relevant documents were retrieved).

The results showed that the controlled language was much superior to any of the others. Of the remainder the free-indexing was superior, followed by the combination of subject headings and modifier line, then title, and title-and-abstract.

The title-and-abstract was shown to have the highest exhaustivity and to give the highest Recall of any of the languages, but it had easily the worst Precision. In other words it has sufficient terms to provide many of the required matches, but many other words which produced matches with non-relevant documents.

The low exhaustivity of the title gave a consistently low Recall but conversely a very high Precision, superior to any of the other languages. Thus its use could be recommended for situations in which only a proportion of all the relevant documents, with as few as possible non-relevant, are required.

The exhaustivity of the combination of the subject heading and modifier line of the abstracts journals was found to be only slightly better than that of the title while the Precision was slightly worse. Thus its overall performance was only a little superior to that of the title.

The free-indexing was shown to have an exhaustivity only slightly worse than that of title-and-abstract and to have a similar maximum Recall, while its Precision was little worse than the subject-heading/modifier combination. It was therefore superior to the other three uncontrolled languages and only outranked by the controlled language.

Probably the most important outcome of the evaluation came from a comparison of the free-indexing and controlled language. Since the document indexing based on these languages contained the same concepts (the controlled-language indexing being the translation of the free-indexing into controlled terms), it followed that the superiority of the controlled language must be attributed to the question formulation. The failure analysis showed that the superiority arises from the built-in ability of the controlled language to include in the formulation all the alternative words and concepts which might be used to express the question.

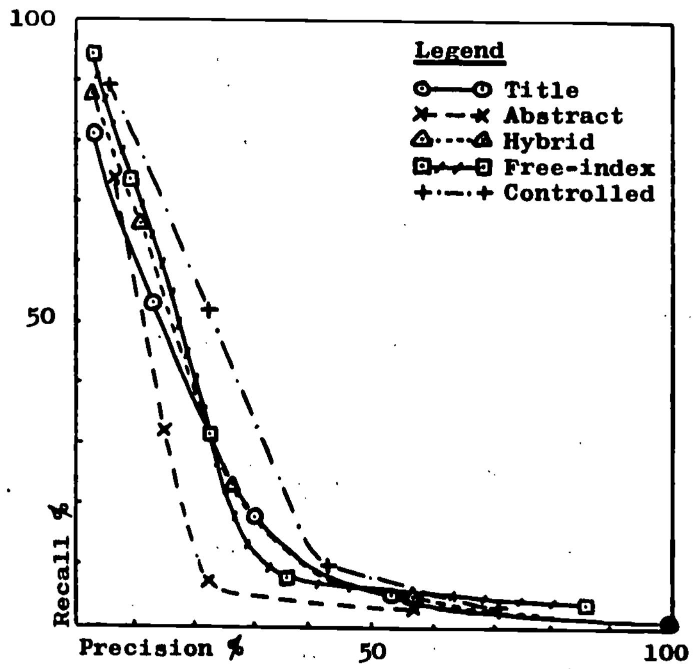


Figure 6

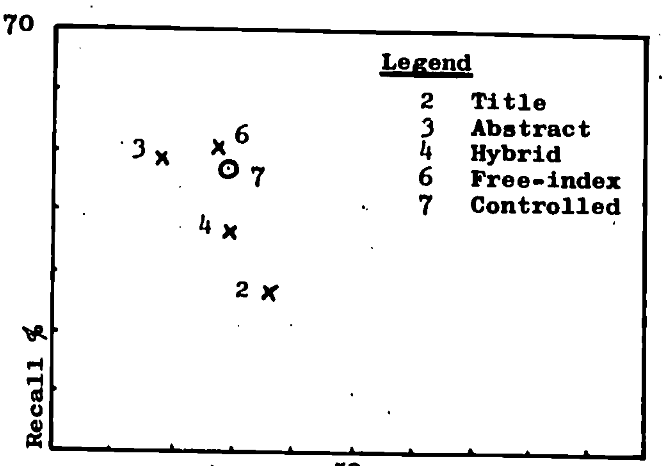


Figure 7

Retrieval Effectiveness

This is a function of the thesaurus and in particular of the procedure of identifying new terms as they occur in the indexing of the input documents and adding them to the thesaurus in their correct relationship to the other terms. This suggests that a similar retrieval performance might be obtained with free-language indexing by the use of a free-language thesaurus for assistance in question formulation. Each new term as it occurred in the document indexing would be added and its relationship to the other thesaurus terms indicated.

It is on these grounds that the decision was made to adopt free-indexing as the basic medium for information retrieval and SDI in the INSPEC system. At present, subject headings and modifier lines are still used in the printed indexes to the abstracts journals, but, as described in the paper by Field in Area , the subject headings are being related to the free-index terms in a thesaurus structure and this will be used as the basis of the free-language thesaurus.

Although in setting up the evaluation project it was not intended to seek results that would necessarily be applicable to other than the INSPEC environment, it is believed that the main conclusions - the superiority of the controlled language and the expectation of obtaining a similar performance by the combination of free-indexing and a free-language thesaurus - are generally applicable.

1. Cleverdon, C., Mills, J., and Keen, M.; Factors determining the performance of indexing systems, vol.1: design; vol.2: test results; Cranfield; Aslib Cranfield Research Project, 1966.

2. Aitchison, T.M., and Tracy, Jennifer M.; Comparative evaluation of index languages. Part I: design. Report No.R.70/1. London, INSPEC, The Institution of Electrical Engineers, July 1969.

3. Aitchison, T.M., Hall, Angela M., Lavelle, Katherine H., and Tracy, Jennifer M.; Comparative evaluation of index languages. Part II: results. Report No.R.70/2. London, INSPEC, The Institution of Electrical Engineers, July 1970.

A RESEARCH TECHNIQUE TO COMPARE THE RETRIEVAL EFFECTIVENESS OF ABSTRACTS AND INDEXING TERMS

T. Bloch, U. Bloch and K.D. Ofer

SYNOPSIS

The method proposed to compare retrieval effectiveness of descriptors and abstracts, endeavours to minimize subjective decisions by duplicating the basic function of extracting the terms from the abstract and by restricting the subject specialist to a few prescribed decisions. The comparison between the two surrogates is based on comparing the number of terms in the various categories, but does not allow for weighing different terms as to their possible effectiveness in retrieval.

The Problem

When using tape services as the input for a mechanized information service, one cannot simply merge the different services together, not only becsuse their formatting is not standardized, but also becsuse the document surrogates vary from service to service, so that the search profiles must be prepared separately for each.

Some services give (in addition to the citation, the author and the title), an abstract and others a set of indexing terms. Several tape services include both surrogates.

In this last case, it is of interest to know, whether one of them, the title together with the abstract, or the set of indexing terms - would give as good a retrieval on its own as together with the other. In other words, whether the set of terms available from the descriptors ("D") and the set of terms from the abstract ("A") are completely overlapping, partly overlapping, or completely disjoint. It would also be of interest to know, which set usually comprises more specific terms than the other.

Once we know the answer to these questions, it should facilitate the design of a better search strategy, e.g. if we know that the terms in the abstract are more specific than the descriptors, we should design the mechanized search program to look first at the descriptors and only after a general term is found in "D" would we bother to search "A" for the more specific term. This could lead us to a two-stage search, where the "A" terms would make for higher precision but the "D"-search would still be available for empty searches. Alternatively, looking only at "D" terms would make for faster throughput of a serial file.

Naturally, these decisions will not be the same for all tape services, but may change from one to the other depending on the quality of the abstract and of the indexing terms. This study shows a research design that could be employed routinely to decide upon the above problem applied to a specific case.

Definitions

1. Equal or synonymous terms are called "joint" terms.
   Expressions are "synonymous", if we can assume, that an information scientist would use either in the construction of the same profile.

2. Where one term is more specific than the other, the two terms are called "inclusive" terms.

3. Any term which is neither joint nor inclusive is a "unique" term.

4. The set of all terms used as descriptors for a document is its "set of descriptor terms" or "D".

5. The set of all terms appearing in the list culled form the title, the bibliographic citation and the abstract of a document is called its "set of abstract terms" or "A".

The Method

The decisions are based on comparisons between the number of terms in the two sets. The number of terms in "D" is usually not difficult to ascertain, but defining the set "A" is more laborious. One cannot simply count the words left over after eliminating all words from a copious "stop list", but one must ensure that only "semantic" words are chosen.

The following technique is suggested:
Abstracts are chosen randomly from the service for which the test is performed. These are photographed three times. From two batches indexing terms are cut away and the remaining title + citation + abstract are given to two intelligent persons, with a good command of English, but with no subject knowledge. They go over the abstracts and list all expressions found in them that are - in their opinion - of sufficient value to rate as indexing terms.

Compatibility between the "indexers" is developed as follows: At the first session the two meet with a subject specialist. After the general ideas of the proceedings are explained, each goes over an abstract on his own and draws up a list of his "set of abstract terms". The resulting two lists are compared, and where they are not identical the differences are discussed with a view to develop rules how to create "A". Wherever possible, these rules are formalized (see Appendix A), In this way the indexing of about a dozen abstracts is practiced.

At the end of the session, each "indexer" takes home another batch of abstracts from which he prepares, on his own, a set of abstract terms, according to the rules. At the next session these two proposed sets are compared and differences are again discussed with a view to adapt the rules. These sessions are repeated until the two lists - prepared at home - are reasonably close.

At this stage another three batches are prepared containing $I + \Delta I$ different abstracts. Again two batches are given to the two participants, but this time no discussion takes place. Instead, the two sets of terms are compared by the subject specialist (who so far had no access to the set of descriptor terms "D"). Where the two lists differ by more than one or two terms, the abstract is excluded from further analysis (ΔI items). Where the lists differ by only one or two terms, the subject specialist may, at his discretion, adjust the two lists to obtain "A", the final set of abstract terms (I items).

Only now, after "A" has been extracted by the above procedure, does the subject specialist compare it to the set of descriptor terms "D". He decides to which of the three classes (joint, inclusive or unique) each term belongs, and which are the more specific ones. This method should provide maximum objectivity on the part of the subject specialist.

In his decision-making process he has merely to compare two lists. He will look first for those terms that appear in both lists and will "make-up" the joint list. He will then have to use his discretion to decide which are synonymous expressions and which are inclusive ones. Finally he goes over the remainder of each list and checks whether these are indeed terms for the "unique" list.

This process was relaxed for only one special case. It happened a number of times, that the two "indexers" who prepared the original lists of "A"; in their zeal to include only "semantic terms" in their lists, both decided to exclude an expression that appeared in the abstract. But when "A" was compared to "D" it turned out that this very same expression (or an obvious synonym) was indeed included in "D". Where this happened, the subject specialist was permitted to add the term to "A". We decided thus because we want principally to compare mechanized retrieval efficiency, and this "lost term" in "A" would certainly have been found by a computer working in a free-text mode.

The next step consists simply of counting, for each item, the number of joint terms ($a_j = d_j = n_j$); of more specific and less specific inclusive terms ($a_{ms} = d_{\ell s}$; $a_{\ell s} = d_{ms}$) and of unique terms (a_u & d_u) for both the set "A" and the set "D". Each of the above counts is summed for all items, thus:

$$n_j = \sum_I a_j \qquad \text{(all joint terms)}$$

$$n_i = \sum_I a_{ms} + d_{ms} \qquad \text{(all inclusive terms)}$$

$$n_{ua} = \sum_I a_u \qquad \text{(all terms unique to the abstract)}$$

V Evaluation of Retrieval Effectiveness

$n_{ud} = \sum_{I} d_u$ (all terms unique to the descriptors)

$n_u = n_{ua} + n_{ud}$ (all unique terms)

$N = n_j + n_i + n_u$ (number of terms given to the batch of size I)

We are now ready to try and answer the original questions: Is there a complete overlap between the two surrogates or does one give appreciably more (unique) terms than the other? Does one of them usually have the more specific terms? How large a recall loss (fewer entry points) will be sustained if only one surrogate is used for retrieval purposes?

Most decisions can be based on descriptive statistics, and no advanced statistical techniques are required. If, for instance, the number of more specific terms in the abstracts is appreciably greater than the number of more specific terms among the descriptors ($\sum a_{ms} >> \sum d_{ms}$), then "A" is the more specific set. If $\frac{n_{ua}}{N}$ is small, we know what fraction of the entry points are lost, should we decide to use only the descriptors. The decision whether to use one or both surrogates, will thus be based mainly on the expected recall loss. A test of significance will only be necessary to make sure that we discard the right surrogate, though in practice this decision will probably be influenced by factors outside this study. To present the data to the system designer, with whom the final decisions lie, it is thus sufficient to turn it into percentages. Confidence levels of the differences can be computed using a non parametric test, such as chi-square or - for very small samples - Fisher's exact probability test.

Example

The above procedure was used to test the surrogates of the U.S.G.R.D.R. (now G.R.A.) service. To check to what extend our findings apply to the various sections, we tested two widely different subjects, namely section 5 covering the behavioural sciences, and section 9 on electronics. That the abstracts of the two sections differ markedly became clear already during the "compatibility phase". Whereas for the behavioural sciences quite a few sessions were required, electronics just one session. As the electronics section was prepared by the same two participants only after they had acquired proficiency, this alone might be attributable to the learning process. But batches in behavioural sciences prepared later were still more discrepant than those prepared for electronics. In the final batches $\frac{\Delta I}{\Delta I + I}$ was 26% for the behavioural sciences and only 5.6% for electronics. This difference computed according to X^{2-} is significant with $P < 0.01$.

Table 1: Comparison of number of terms in "A" & "D" in the behavioural sciences.

	No.	%
I	37	
N	624	100
n_j	139	22.2
n_i	65	10.6
n_{ua}	170	27.2
n_{ud}	250	40
a_{ms}	54	8.8
d_{ms}	11	1.8

Table 2: Comparison of number of terms in "A" & "D" in electronics.

	No.	%
I	35	
N	458	100
n_j	104	22.7
n_i	35	7.7
n_{ua}	194	42.3
n_{ud}	125	27.4
a_{ms}	25	5.4
d_{ms}	10	2.2

V Evaluation of Retrieval Effectiveness

The various counts for these final two batches are given in Table 1 (for behavioural sciences) and Table 2 (for electronics). Clearly, there is little overlap between the two surrogates, as fully 2/3 of the terms are unique. It is perhaps of interest to point out that, while in one section one surrogate provides more unique terms, in the other the position is exactly reversed. Thus we cannot say which surrogate may be dropped (should one really be prepared to forgo a quarter of the entry points).

The more specific of the inclusive terms are always more numerous within the abstracts with a significance level of $0.01 < p < 0.02$ for section 9 (electronics); and of $p < 0.01$ for both section 5 (behavioural sciences) and the two combined sections. It appears that the process discribed in the beginning, namely to use "D" as a sieve for pinpointing possible hits and use the abstract for the final check, whether the item is relevant, may be useful (although the many unique terms in the abstract make this technique less attractive).

Before taking any final decision we would cross-validate the results, by repeating the experiment - possibly on another two sections.

Appendix A

Examples of "Rules how to prepare set "A".

1. List all words you deem relevant and all words you do not understand.
2. If synonyms appear, list them.
3. Keep concepts separate.
4. List all countries except U.S.A.
5. Abbreviations must be written out.
6. Look at the bibliographic details to see if they add anything relevant.
7. Formulas and numerical values shall not be listed.

Acknowledgements

We are grateful to Mr. E. Gross, who originally suggested the subject of comparing different surrogates and proofread the manuscript. We are especially grateful to Miss H. Ben-Joseph and Mrs. G. Haimovic who prepared the sets of abstracts terms and spent many an evening in "becoming compatible".

EVALUATION OF RETRIEVAL EFFECTIVENESS IN RELATION TO USER REQUIREMENTS.
SOME CASE STUDIES FROM THE COMPUTERIZED INFORMATION RETRIEVAL SYSTEM AT THE ROYAL INSTITUTE OF TECHNOLOGY, STOCKHOLM.

Z. Gluchowicz
Head of Documentation Department
Royal Institute of Technology Library, Stockholm

SYNOPSIS

The report gives a short description of the computerized information retrieval system -SDI- (Selective Dissemination of Information) at the Institute. The user population is presented and the distribution of profiles by subject categories is indicated.
The evaluation of the retrieved information is based on the feedback from users according to the Katter scale.
The cases studied represent real requests made to the system by types of users representing research at the industrial community and at universities. The requests chosen represent science-oriented and discipline-oriented queries. The SDI system at the Institute has been in operation since 1967. About 1200 queries generating about 2000 profiles have been run in the system during periods of varying length from six months to three years.

V Evaluation of Retrieval Effectiveness

Introduction.
Evaluation of effectiveness for information retrieval systems is frequently expressed in terms of precision and recall ratio although there is a variety of components influencing the effectiveness of information search.
The precision ratio is generally described as the proportion of relevant items retrieved over the total items retrieved and the recall ratio is the proportion of relevant items retrieved over the total relevant items in the system.
Since in real-user environment the user of an operating retrieval information system is the only person to make evaluation of relevance, precision and recall depend on subjective estimation of relevance.
Precision figures based on evaluation relevance of real requests made to an information system may be established with a reliable degree of accuracy, while recall figures for a big information system can only be estimated hypothetically.
The present paper deals with retrieval effectiveness evaluated by users in terms of relevance. The precision figures are based on the performance of real requests submitted to an operative information retrieval system.

System description.
The Royal Institute of Technology, Stockholm has developed a computerized information retrieval system -SDI- Selective Dissemination of Information - with the goal to ensure an extensive ready access to documentation in subject fields of interest for the industrial community and for the educational and research program at the universities and research institutes in Sweden. At present the SDI system processes 14 different databases, 11 which are externally generated and 3 which are internally generated. (Table 1) The total amount of references processed exceeds one million references a year.
Each database has its own record format designed and developed by the system originator.
The problem of incompatibility of tape information recorded in different formats has been solved at the Institute by the mean of a general program system for information retrieval - ABACUS. The system has been developed through the initiative and guidance of the head librarian Björn Tell.(5)

Profile performance.
The profile is a list of terms significant to the query which will potentially be found in the references. (Table 2)
Terms or groups of terms are linked together in a logical manner as expressed in the search strategy. The required connections between words or groups of words are expressed by using Boolean logic, the logical operators being "OR", "AND" , "NOT".
The search strategy may be expressed by one or several alternative logical connections, each connection may contain several parameters.
The symbols for the logical operators used at the Institute are:
"|" for "OR", "&" for "AND", "¬" for "NOT".
Profile 87 B (Table 2) is presented here as an illustration of a search strategy comprising four alternative logical connections. Each connection contains one or several parameters which the items retrieved have to satisfy.

Name of Tape	Abbreviation	Editorial Organisation	Scientific field	Type of Document analysed	Contents of Document	Document/yr
Source Data Tape		Institute for Scientific Information	Science, technology, medicine	2200 journals	Title words, authors, author's address, journal title	350,000 (1969)
KTH Mechanical Engineering	KTH-MechEng	Royal Institute of Technology Library	Mechanical engineering, metallurgy	200 journals	Expanded title, author, journal title.	60,000 (1970)
Polymer Science & Technology Patents	POST-P	Chemical Abstracts Service	Polymer-chemistry, plastics technology	Patents	Title words, author journal title, digest	14,000 (1970)
Chemical Abstracts Condensates	CAC (even no:s)		Chemistry	600 journals	Title words, author journal coden	130,000 (1969)
INSPEC Tape Services		Institution of Electrical Engineers	Physics, electrical, electronic, computing and control engineering	1600 journals, reports, patents, conferences, books	Titles, subject headings, language, author, institutions, free terms, subject class, journal title	120,000 (1970)
Metals Abstracts Index Data Base	METADEX	American Society for Metals	Metallurgy, relevant physics and chemistry	1000 journals monographs theses	Titles, authors, keywords, journal titles	25,000 (1970)
Current index to Conference Papers		CCM Information Corp.	Chemistry Engineering Life Science	Meetings, papers	Name of meeting, title of paper, authors, thesaural terms, free terms, meeting subject.	300,000 (1970)

Table 1

Name of Tape	Abbreviation	Editorial Organisation	Scientific field	Type of Document analysed	Contents of Document	Document/yr
Computerized Engineering Index	COMPENDEX	Engineering Index, Inc.	Engineering some applied science and management	3500 journals, reports etc.	Index terms, free language terms, authors, journal titles.	70,000 (1968)
Nuclear Science Abstracts	NSA	US Atomic Energy Commission	Nuclear science and technology	2000 journals	Classification, titles, authors subject, report no.	47,000 (1967)
Citation tape for the abstract bulletin of the Institute of Paper Chemistry		Institute of Paper Chemistry	Pulp and paper science and technology	800 journals, patents, etc.	Abstract no:s, authors, journals, dates, patent no:s	11,000 (1970)
Food Science and Technology Abstract.		International Food Information service	Food science and chemistry	Journals, patents etc.	Title, author, abstract no, patent no, journal, affiliation, descriptor.	12,000
Educational Resources in Information Center	ERIC	International Resources Information Center	Educational science, pedagogics	Reports, journals.	Keyword, title words, author, journal, report no.	29,000
KTH-Wood		Royal Institute of Technology Library	Wood cutting, handling, working	40 journals	Expanded title, author, journal title	6,000 (estimated amount 1971)
Accession list from KTH, CTH, Ae	Nyförvärvs-lista	"	Science	Recently acquired books, reports, conferences, dissertations, sp. issues of journals.	Author, title, printing year, report no, library	6,300 (1970)

Table 1

Profile 87 B

Subject field: Array antenna design.

Search strategy: ¬E&(A&(2D|C&D)| B&(C|D))

Nr.	Grp	Search term
0010	A	ANTENNA*
0020	A	APERTUR*
0030	A	AERIAL*
0040	B	ARRAY*
0050	C	PHASE COMPARISON*
0060	C	MONOPULS*
0070	D	EQUIVALEN*
0080	D	PATTERN*
0090	D	SYNTHES*
0100	D	EXCITAT*
0110	D	DIFFEREN*
0120	D	DESIGN*
0130	D	SIDELOBE*
0140	D	DISTRIBUT*
0150	D	PLANAR*
0160	D	GAIN*
0170	D	TAPER*
0180	D	RADIAT*
0190	D	COLLINEAR*
0200	D	COHERENT*
0210	D	DIRECTIVIT*
0220	D	SIDE LOPE*
0230	D	SLOPE*
0240	D	ERROR*
0250	D	CIRCULAR*
0260	E	ENDFIRE*
0270	E	YAGI
0280	E	CYLINDRIC*

* truncation flag

Profile printout

Table 2

V Evaluation of Retrieval Effectiveness

According to search strategy $\overline{E}\&(A\&(2D|C\&D)|B\&(C|D))$ an item will be retrieved if any one of four following connections is satisfied:

1. A&2D
2. A&C&D
3. B&C
4. B&D

and none of the terms in group E appears in the references.

Any one of the connections 1-4 will be satisfied if the item retrieved contains one of the following combinations:

Connection 1: (word 10 OR word 20 OR word 30) AND (word 70 AND word 80 OR word 70 AND word 90 OR ... OR word 70 AND word 250 OR word 80 AND word 90 OR word 80 AND word 100 OR ...OR word 240 AND word 250) but NOT (word 260 OR word 270 OR word 280)

Connection 2: (word 10 OR word 20 OR word 30) AND (word 50 OR word 60) AND (word 70 OR word 80 OR ...OR word 250) but NOT (word 260 OR word 270 OR word 280)

Connection 3: word 40 AND (word 50 OR word 60) but NOT (word 260 OR word 270 OR word 280)

Connection 4: word 40 AND (word 70 OR word 80 OR ...OR word 250) but NOT (word 260 OR word 270 OR word 280)

These connections represent 572 alternative possibilities for items to be retrieved.

Profiles distribution (June 1969).

User groups:

Government and local administration	40
Government research institutes	106
Trade associations and institutes	17
Universities	193
Libraries	7
Private companies	478
Individual researches	4
	845

Subject categories:

Mechanical engineering	179
Iron, Steel and metal manufacturing	133
Polymer technology	80
Chemical technology	77
Physics	64
Electrotechnology	59
Control and computer technology	57
Reactor technology	40
Wood, pulp and paper	38
Telecommunication	34
Transportation, packaging	27
Business administration	27
Inventories and storehousing	15
Food industry	15
	845

User-system interaction.

In a batch processing information system the user-system interaction is maintained by means of a dialogue between user and co-ordinator. When submitting a query to the SDI system the user has to provide a statement of his fields of interest in natural language in a narrative way giving a detailed description of his subject field. He may also provide a draft proposal for his search profile. This he can do on the basis of the Profile Design Manual issued by the system author.

The first stage of a user-system interaction is the dialogue between the user and the documentalist at the centre clarifying the scope and nature of the query. As the user is the only person to give a correct statement of the requests the statement should preferably be made by himself and not by an intermedary such as a librarian or an information officer. Therefore the documentalist at the SDI centre insists on maintaining direct contact with the information user.(Table 3)

The second stage of the user-system interaction takes place when the profile is being defined. This dialogue is necessary for selecting search terms for the profile and for the decision on the search strategy.

GENERALIZED FLOW CHART FOR PROFILE CONSTRUCTION

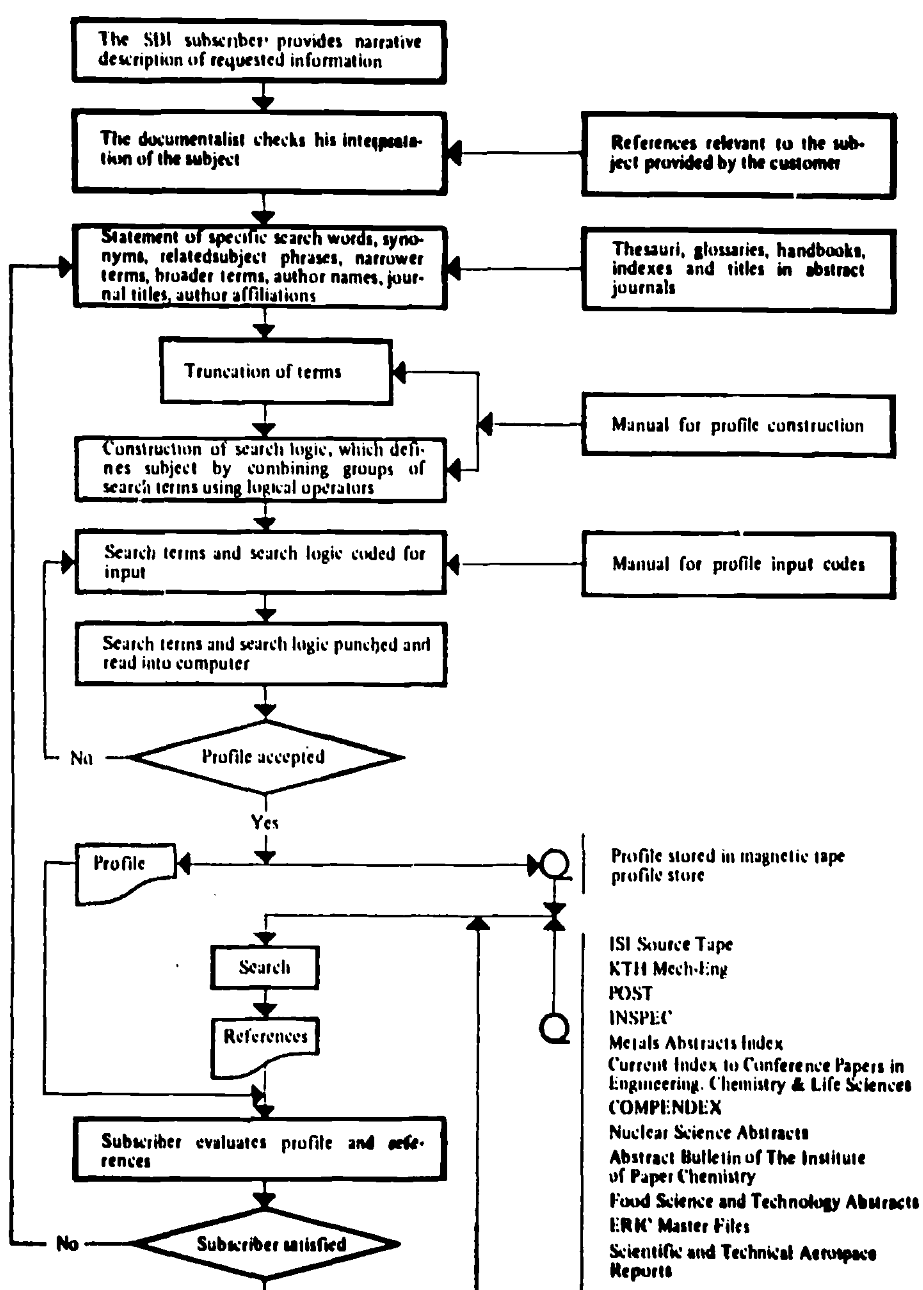


Table 3

Selektiv Delgivning av Information
Kungl Tekniska Högskolans Bibliotek
Dokumentationstjänst

LITTERATURLISTA

datum 29/12/70

Formulär 3

beställare

kontaktperson

företag/institution

sökprofil nr 87B

adress

postadress

ARRAY/ GAIN/ 1
BORGIOTTI, G.V.
ALTA FREQUENZA (ITALY)
VOL.39, NO.8 701-205 AUG. 1970 12 B7038246
IMPEDANCE AND GAIN OF A DIPOLE IN AN INFINITE PERIODIC PHASED ARRAY

ANTENNA/ PLANAR/ DIRECTIVIT/ 2
FORMAN, B.J.
RADIO SCI. (USA)
VOL.5, NO.7 1077-83 JULY 1970 B7038254
A NOVEL DIRECTIVITY EXPRESSION FOR PLANAR ANTENNA ARRAYS

ARRAY/ RADIAT/ 3
KILLICK, E.A. ZOLEDZIOWSKI, K.A. PORTER, N.E.
1969 EUROPEAN MICROWAVE CONFERENCE
399-402 1970 SEP 1969 B7038257
RADIATION OF DUAL POLARISATION FROM LINEAR ARRAYS

ARRAY/ DESIGN/ 4
HART, G. VIOLA, R.
1969 EUROPEAN MICROWAVE CONFERENCE
443-6 1970 SEP 1969 B7038258
DESIGN OF L, C AND S BAND SOLID STATE ARRAY MODULES

ARRAY/ DESIGN/ 5
PATTON, W.T STAIMAN, D.
1970 IEEE INTERNATIONAL CONVENTION DIGEST
390-1 1970 MAR 1970 B7038259
INTEGRATED CIRCUITS IN ARRAY DESIGN

ARRAY/ DESIGN/ 6
ROSENBLATT, A.
ELECTRONICS (USA)
VOL.43, NO.15 78-81 20 JULY 1970 B7038821
VERSATILITY IS DESIGNED INTO DUAL BAND MODULE FOR PHASED ARRAY SYSTEMS

Table 5

V *Evaluation of Retrieval Effectiveness*

On the basis of the user's statement the documentalist at the centre compiles the list of search terms which are significant to the query and are potential terms in the references. As the SDI system at the Institute is a free-language system the exhaustiveness of the terms in the profile is an important parameter in retrieval efficiency.
Submitting a query to the system the subscriber may indicate on the order card his requirements for precision and the search strategy is formulated according to these indications.

The search output - list of retrieved references - (Table 5) is delivered to the subscriber together with an evaluation form designed according to Katter's six grade scale. (Table 4)
During the experimental period the SDI service at the Institute was free of charge and the user was asked to evaluate the information received. The users' evaluation was used as feedback for modifying the profiles in order to obtain precision and recall corresponding to the users' needs and judgment.
At the feedback stage the documentalist at the centre amends the search strategy and/or search terms in order to obtain greater search efficiency.
Profile adjustment continues the dialogue between the user and documentalist.

During the fiscal year 1968-69 a total of 36072 references were evaluated. These references have been selected by the SDI system on the basis of individual profiles, which as of June 30 1969 numbered 845.

Users' evaluation	References	Percentage
1. Of immediate interest	9080	25.1
2. Of interest, I have already read it	2459	6.8
3. Of interest, but not for immediate use	11009	30.6
4. Cannot determine interest, because the citation does not provide enough detail	1306	3.6
5. Of no interest, because material does not correspond to what I have described to the system.	11913	33.0
6. Of no interest, because my interests have changed since I described them to the system.	305	0.9
	36072	100.0

Table 4

The figures on Table 4 are based on relevance evaluation of real requests in real user environment during the free-of-charge period.

Experiences.
When the SDI service at the Institute became subject to an annual charge ($60: a year) several subscribers continued to evaluate the received information. They appreciated these feedback possibilities for amending profiles.
The users were aware of two significant factors:

A) The SDI system at the Institute is not a static system, development work is continously done with emphasis on programming modifications and system techniques. New databases are being included in the system in order to obtain more comprehensive coverage in the subject fields of the queries.

B) The users' fields of interest is not static either. The priority of certain aspects of the subjects of interest may change at various times and the profiles have to be modified corresponding to these changes. The staff of the system must often review the profiles and suggest appropriate changes. The documentalists' initiative and suggestions are appreciated especially by these users who seldom do anything to their profiles.

The great majority of users who entered their subscription to the system after the free-of-charge period are evaluating the retrieved information during the initial period when the profile has to be amended frequently. During this period the staff at the centre is analysing the failures such as wrong correlation of terms, overrestrictiveness or broadness of search strategy etc. When the search results are satisfactory to the user's need, the user often ceases to make evaluation assessments.

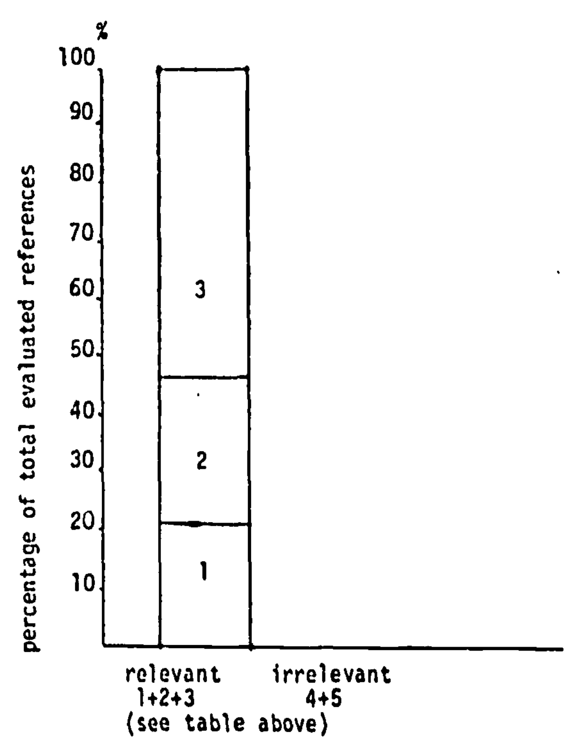
The cases illustrated in this paper were chosen because of the longest duration of the evaluation period. They are cited according to the chronological sequence that the queries were submitted to the system. Of the 15 cases illustrated, 7 stem from private industrial firms, marked with the letter "I", 5 stem from governmental research institutes, marked with "G"and 3from Universities marked with "U". The databases searched in each case are also listed according to the chronological sequence in which they have been searched. As the Source Data Tapes have been searched since 1968 the biggest number of references in the illustrated cases come from this database. (Table 6)

Case 1.

Query: Theory of roller bearings. Theory of lubrication. (I)

Evaluation period: 9.12.1968 - 5.2.1971

database		number of references						
			evaluated					
first search	name of tape	retrie-ved	total	1	2	3	4	5
9.12.1968	ISI	271	214	49	49	115	-	1
11.2.1969	MECH-ENG	65	58	10	18	30	-	-
30.1.1970	METADEX	35	29	7	3	19	-	-
16.10.1970	COMPENDEX	27	22	3	12	7	-	-
total		398	323	69	82	171	-	1
percen-tage				21,5	25,5	53,0	-	-



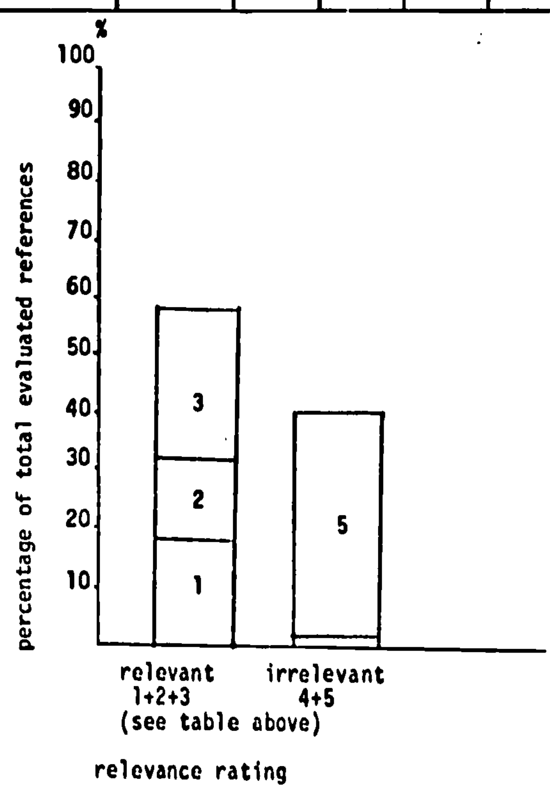
relevance rating

Case 2.

Query: Mobile radio communication equipment (1)

Evaluation period: 11.2.1969 - 3.2.1971

database		number of references						
			evaluated					
first search	name of tape	retrieved	total	1	2	3	4	5
11.2.1969	ISI	32	24	4	4	7	-	9
29.12.1969	INSPEC	50	35	8	4	8	1	14
28.10.1970	COMPENDEX	4	4	-	1	1	-	2
total		85	63	12	9	16	1	25
percentage				19,1	14,3	25,4	1,6	39,6

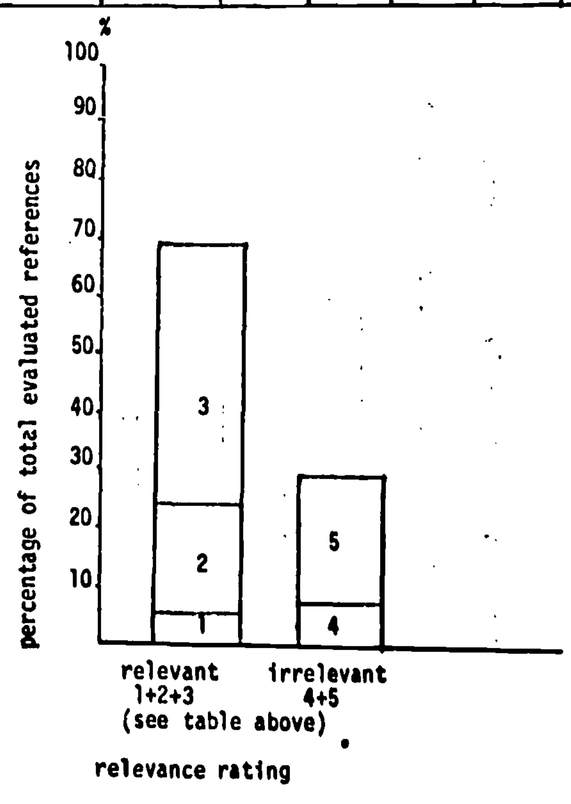


Case 3 ·

Query: Noise generated by hydrodynamic devices (G)

Evaluation period: 21.2.1969 - 3.2.1971

database		number of references							
first search	name of tape	retrieved	evaluated total	1	2	3	4	5
21.2.1969	ISI	84	45	1	8	21	5	10
16.1.1970	INSPEC	70	44	4	9	19	2	10
total		154	89	5	17	40	7	20
percentage				5,6	19,1	45,0	7,9	22,4

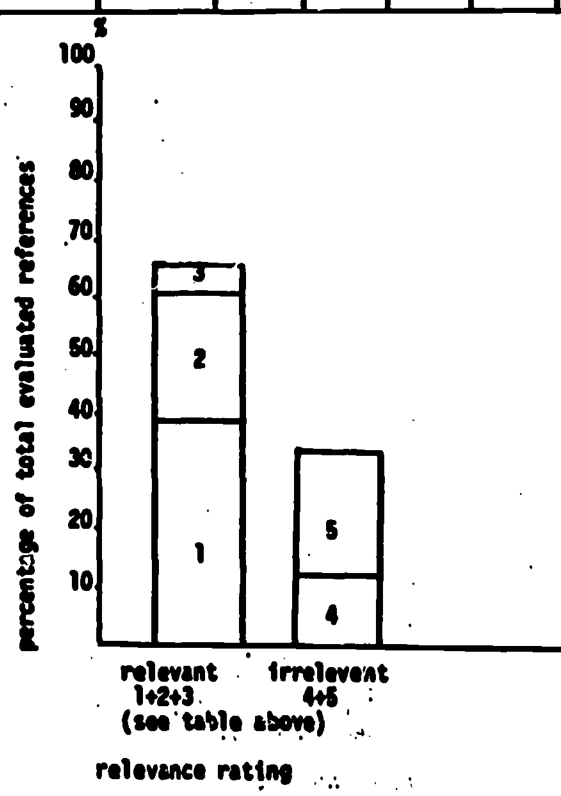


Case 4.

Query: Fluidic devices (G)

Evaluation period: 16.4.1969 - 5.2.1971

database		number of references						
			evaluated					
first search	name of tape	retrieved	total	1	2	3	4	5
15.5.1969	ISI	315	42	20	11	3	3	5
16.4.1969	MECH-ENG	174	44	17	9	3	6	9
18.12.1969	INSPEC	279	153	57	34	3	21	38
total		768	239	94	54	9	30	52
percentage				39,3	22,6	3,8	12,6	21,7

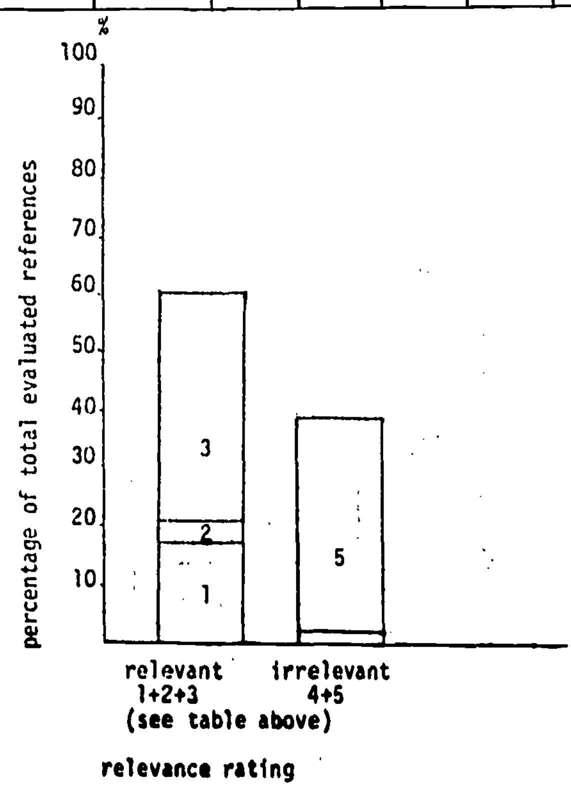


Case 5.

Query: Calculation of electrical power networks (I)

Evaluation period: 2.5.1969 - 5.2.1970

database		number of references							
			evaluated						
first search	name of tape	retrieved	total	1	2	3	4	5
2.5.1969	ISI	290	245	52	13	97	6	77
29.5.1969	MECH-ENG	29	14	-	-	7	-	7
18.12.1969	INSPEC	645	489	80	7	201	8	193
16.10.1969	COMPENDEX	56	28	3	6	5	-	14
total		1020	776	135	26	310	14	291
percentage				17,4	3,3	40,0	1,8	37,5

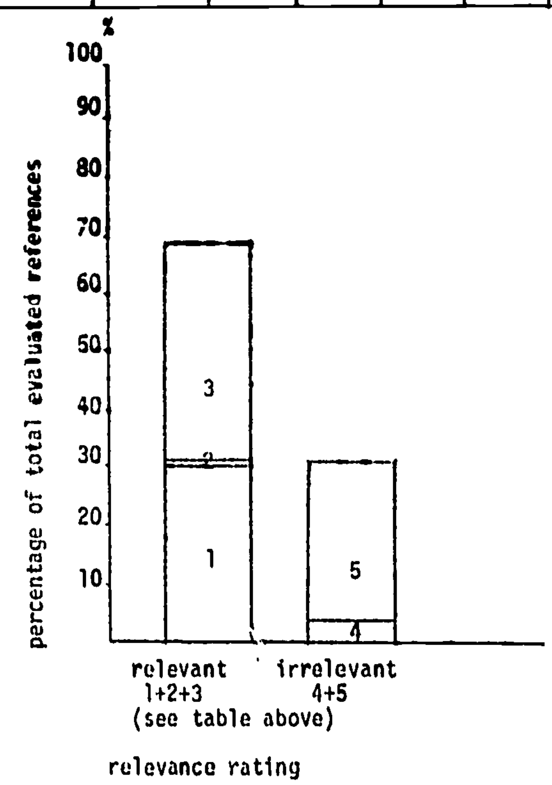


Case 6.

Query: **Cereal-based foodstuffs** (I)

Evaluation period: 2.5.1969 - 21.12.1970

database		number of references							
			evaluated						
first search	name of tape	retrieved	total	1	2	3	4	5
2.5.1969	ISI	504	491	147	3	191	18	132
16.10.1970	COMPENDEX	3	3	-	-	-	-	3
total		507	494	147	3	191	18	135
percentage				29,8	0,6	38,6	3,6	27,4

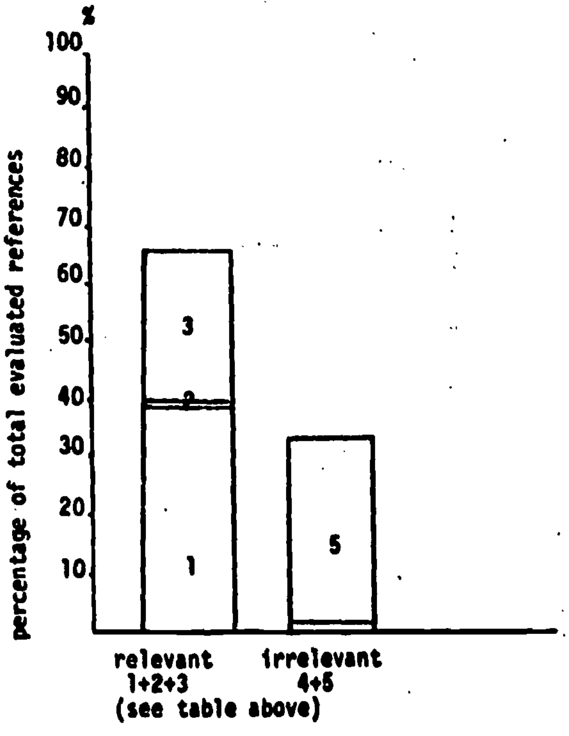


Case 7.

Query: Compressors (I)

Evaluation period: 27.5.1969 - 5.2.1971

database		number of references						
			evaluated					
first search	name of tape	retrieved	total	1	2	3	4	5
30.5.1969	ISI	276	219	50	-	57	4	108
27.5.1969	MECH-ENG	247	232	151	4	52	2	23
28.2.1970	INSPEC	50	50	6	-	14	-	30
16.10.1970	COMPENDEX	50	39	6	-	17	-	16
total		623	540	213	4	140	6	177
percentage				39,5	0,7	25,9	1,1	32,8

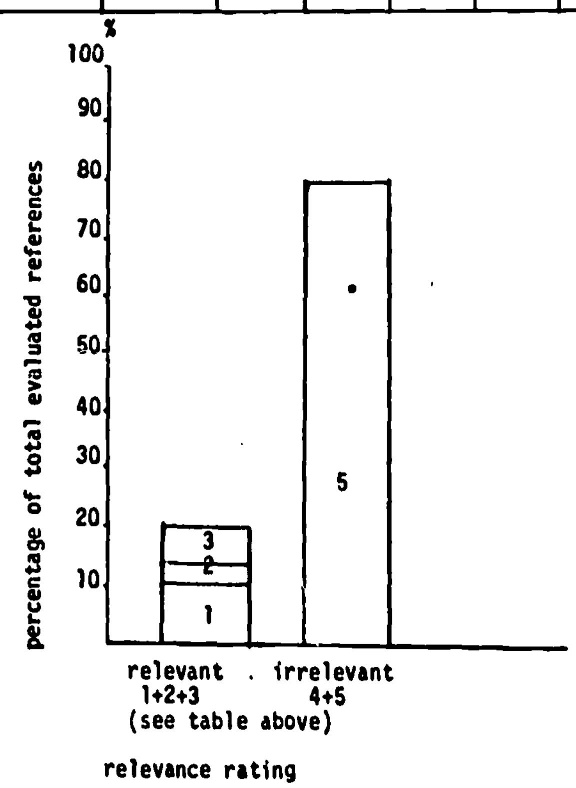


Case 8.

Query: Technical forecasting on the development of electronics (G)

Evaluation period: 13.8.1969 - 5.2.1971

database		number of references						
			evaluated					
first search	name of tape	retrieved	total	1	2	3	4	5
13.8.1969	ISI	206	179	12	6	14	1	146
8.12.1969	INSPEC	427	412	50	14	21	1	326
16.10.1970	COMPENDEX	22	13	1	-	1	-	11
total		655	604	63	20	36	2	483
percentage				10,4	3,3	6,0	0,3	80,0

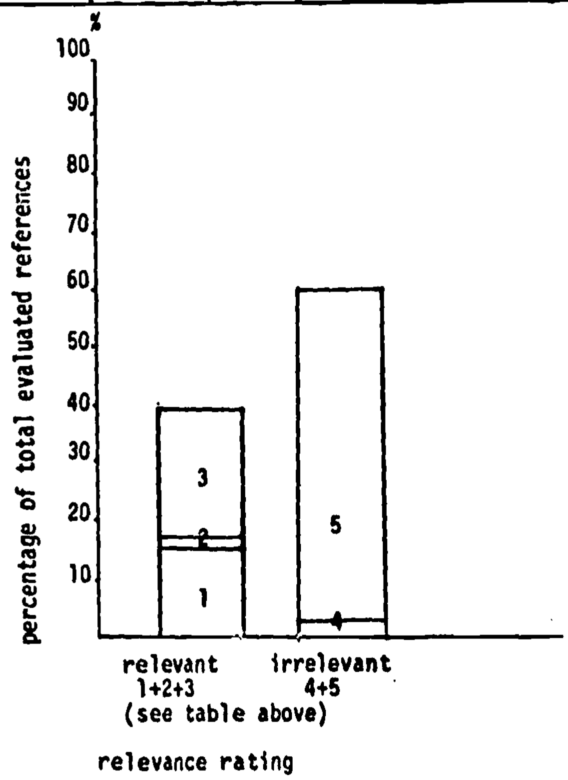


Case 9.

Query: Pumps and pumping (I)

Evaluation period: 19.8.1969 - 5.2.1971

database		number of references						
				evaluated				
first search	name of tape	retrieved	total	1	2	3	4	5
20.8.1969	ISI	435	393	55	12	66	17	243
19.8.1969	MECH-ENG	492	419	63	4	115	5	232
16.10.1970	COMPENDEX	33	17	7	-	5	-	5
total		960	829	125	16	186	22	480
percentage				15,1	1,9	22,4	2,6	58,0

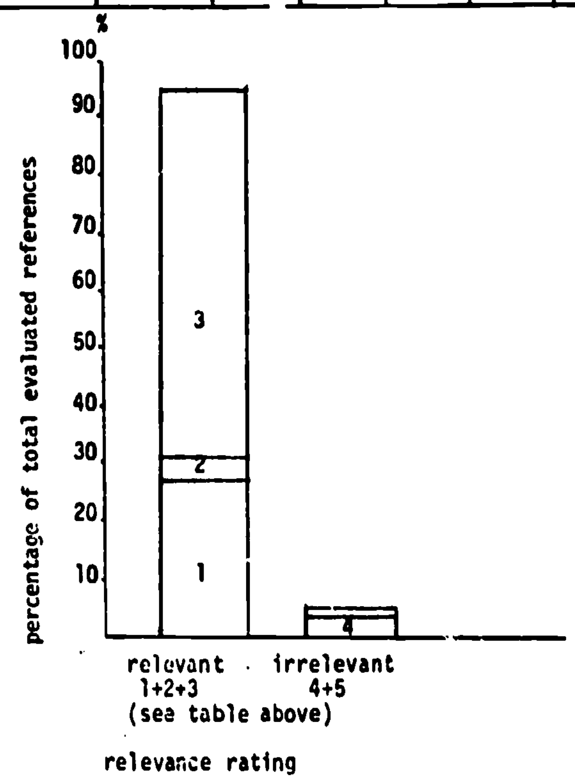


Case 10.

Query: Chlorinated hydrocarbons in different organisms (U)

Evaluation period: 25.8.1969 - 23.1.1971

database		number of references						
			evaluated					
first search	name of tape	retrieved	total	1	2	3	4	5
25.8.1969	ISI	739	685	192	26	460	5	2
8.9.1969	POST-P	56	43	4	-	13	20	6
26.9.1969	POST-J	24	17	5	-	7	5	-
total		819	745	201	26	480	30	8
percentage				27,0	3,5	64,4	4,0	1,1

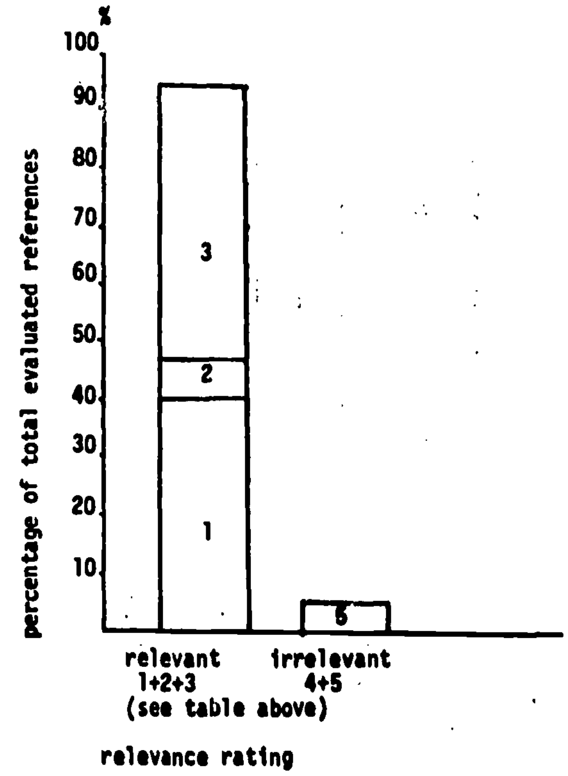


Case 11.

Query: Nickel-iron batteries, iron electrodes (U)

Evaluation period: 3.9.1969 - 5.2.1971

database		number of references						
			evaluated					
first search	name of tape	retrieved	total	1	2	3	4	5
3.9.1969	ISI	50	22	17	2	2	-	1
6.9.1969	MECH-ENG	8	8	2	-	5	-	1
30.1.1970	METADEX	30	26	5	2	18	-	1
9.3.1970	INSPEC	6	5	1	-	4	-	-
total		94	61	25	4	29	-	3
percentage				41,0	6,6	47,5	-	4,9

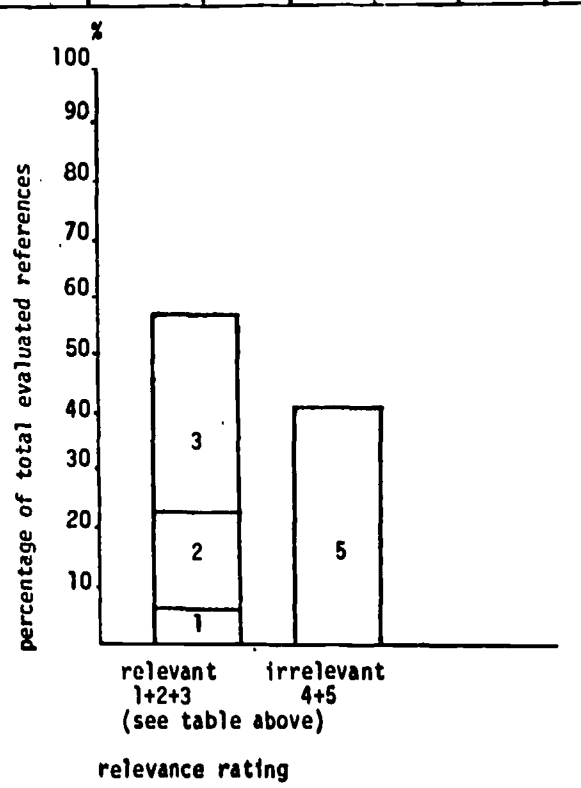


Case 12.

Query: Aerials (G)

Evaluation period: 6.9.1969 - 5.2.1971

database		number of references						
			evaluated					
first search	name of tape	retrie-ved	total	1	2	3	4	5
6.9.1969	ISI	356	332	12	52	112	-	156
29.12.1969	INSPEC	53	40	9	12	18	-	1
16.10.1970	COMPENDEX	6	4	-	1	3	-	-
total		415	376	21	65	133	-	157
percen-tage				5,6	17,3	35,3	-	41,8

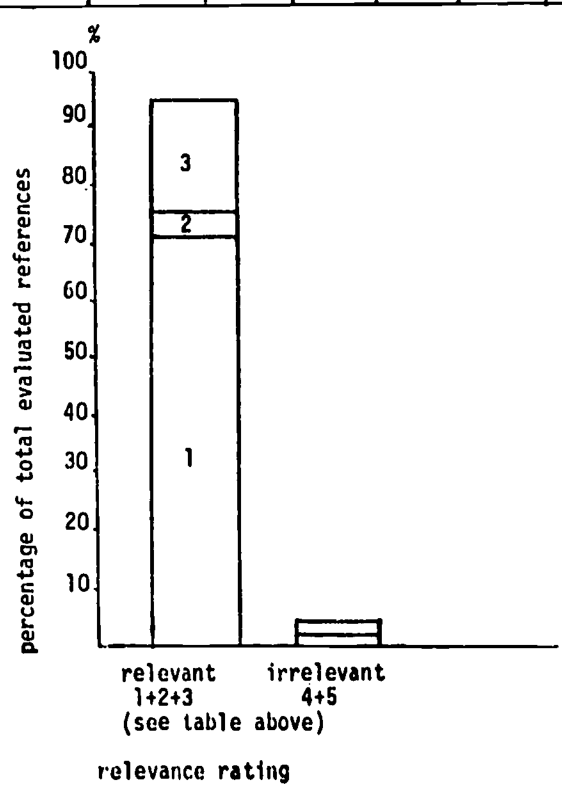


Case 13.

Query: Optimization of preventive maintenance (I)

Evaluation period: 7.11.1969 - 5.2.1971

database		number of references						
first search	name of tape	retrieved	evaluated total	1	2	3	4	5
12.11.1969	ISI	107	99	73	3	17	3	3
7.11.1969	MECH-ENG	79	71	43	6	20	2	-
28.2.1970	INSPEC	88	87	65	3	15	1	3
16.10.1970	COMPENDEX	20	11	10	-	1	-	-
total		294	268	191	12	53	6	6
percentage				71,3	4,5	19,8	2,2	2,2

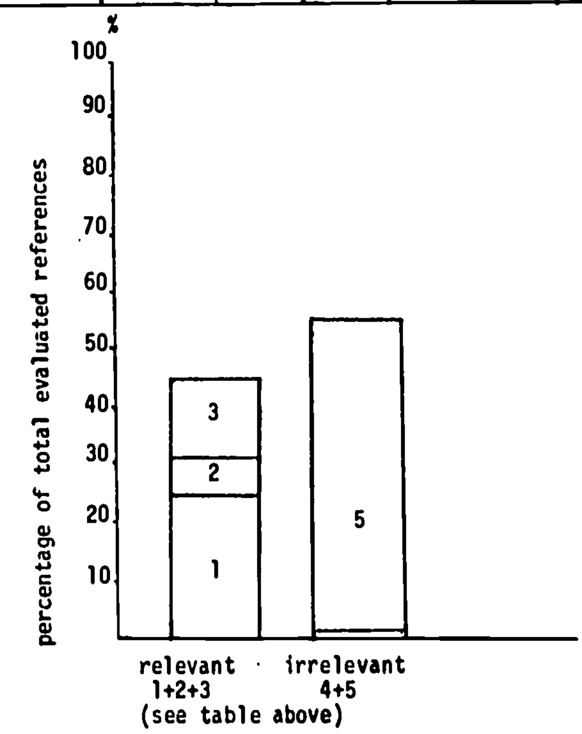


Case 14.

Query: Information on radiation protection (G)

Evaluation period: 30.3.1970 - 23.1.1971

database		number of references						
first search	name of tape	retrieved	evaluated total	1	2	3	4	5
30.3.1970	ISI	695	642	139	47	74	10	372
1.4.1970	INSPEC	281	247	70	23	17	4	133
26.5.1970	NSA	575	436	115	30	83	-	208
total		1551	1325	324	100	174	14	713
percentage				24,4	7,5	13,1	1,1	53,9



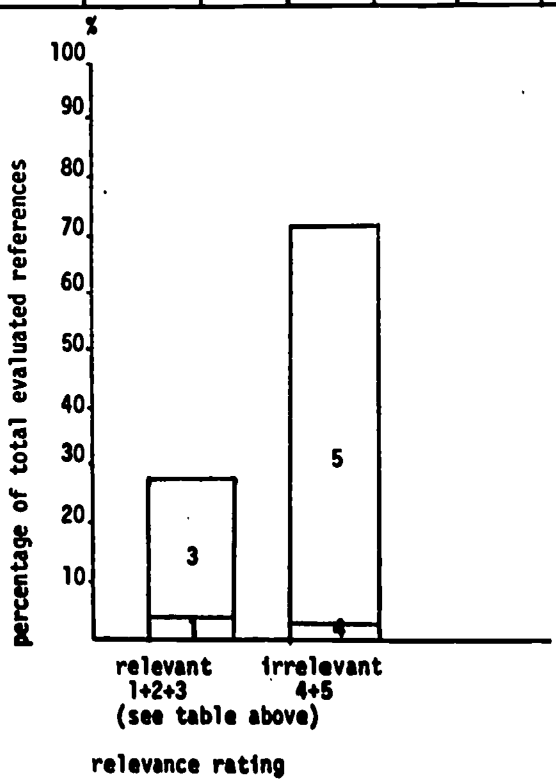
relevance rating

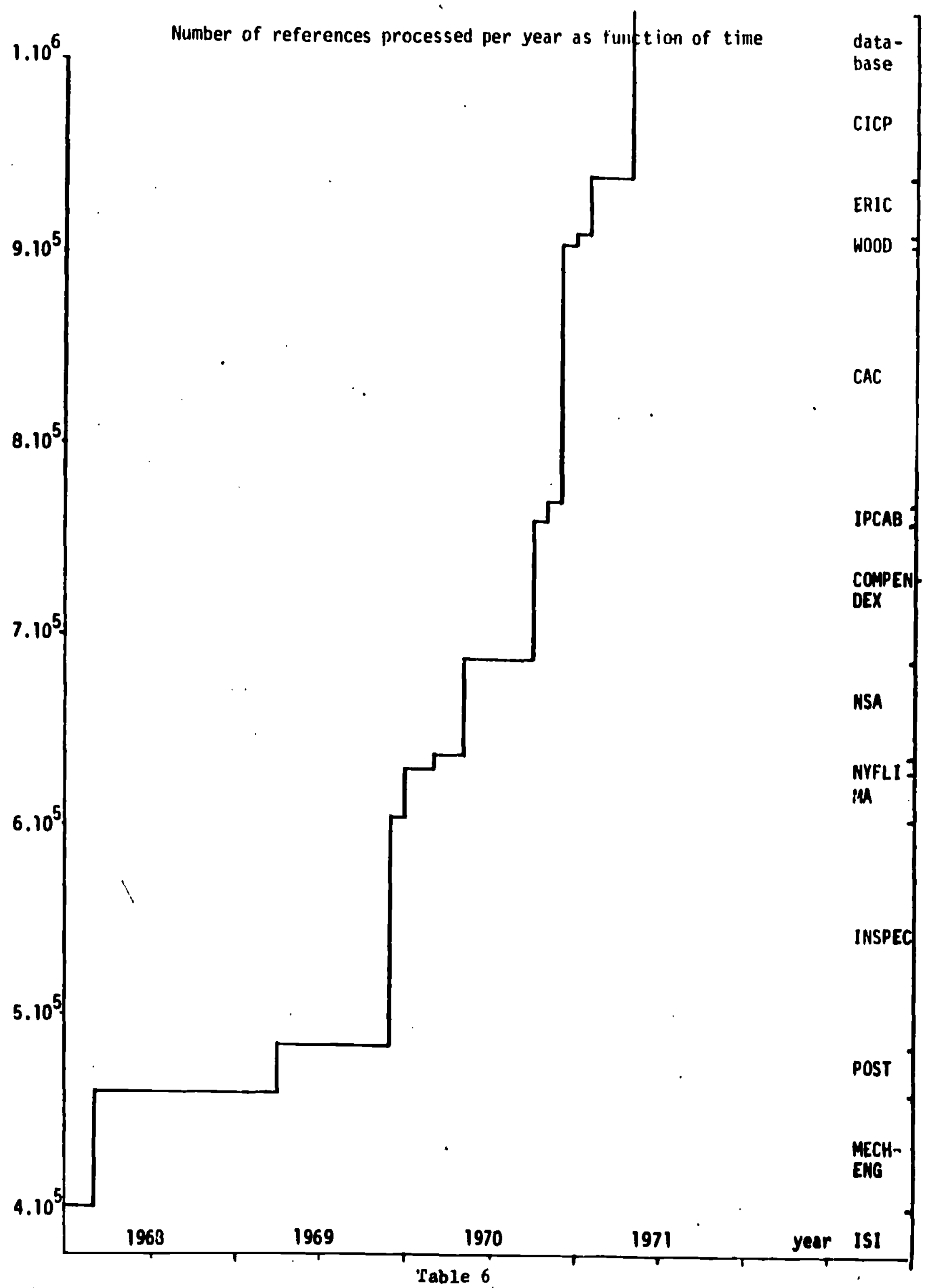
Case 15.

Query: Influence of mineral nutrients on morphogenesis (U)

Evaluation period: 2.6.1970 - 23.1.1971

database		number of references							
			evaluated						
first search	name of tape	retrieved	total	1	2	3	4	5
2.6.1970	ISI	1032	1032	44	1	245	29	713
total		1032	1032	44	1	245	29	713
percentage				4,3	0,1	23,8	2,8	69,0





Number of references processed per year as function of time
1.10^6
9.10^5
8.10^5
7.10^5
6.10^5
5.10^5
4.10^5
1968
1969
1970
1971
year
data-
base
CICP
ERIC
WOOD
CAC
IPCAB
COMPEN
DEX
NSA
NYFLI
MA
INSPEC
POST
MECH-
ENG
ISI

Table 6

Conclusions.
According to Heaps (8) the effectiveness of mechanised information retrieval system expressed in terms of relevance may be represented as a function of question, estimate of relevance and database.

Effectiveness = F (question, estimate of relevance, database)

This functional definition of effectiveness is consistent with our three years of experience with the SDI system at the Institute.
The user's information need, his statement of request, his relevance requirements and relevance judgments, his tolerance of irrelevance makes him a significant parameter of this functional definition of effectiveness. The system exists for the user and in order to satisfy the user's need. The databases have to be chosen with the goal of covering the fields of interest of the queries submitted to the system.
In the SDI system at the Institute most of the profiles are searched on several databases, because the queries are more problem-oriented than discipline-oriented. The distribution of profiles among the databases as of December 1970 is shown in the table below.

Number of profiles	Number of databases searched
62	1
58	2
63	3
201	4
187	5
32	6
4	7

Although the system is based on a multiple data file, there was no significant overlap among the data bases. Infact they tend to be complementary by handling different aspects of subjects.
The effectiveness of the search results in suitable databases has depended in a very high degree on the active interest of the subscriber. The user with experience and training in information work or with some knowledge of the basic principles of a mechanised information retrieval system will have a more meaningful dialogue with the system. He takes a more active part in the development of his profile and his requirements of precision and recall are easier to satisfy because they are more realistic. According to our experience scientists working with basic research are more inclined to use a broader search strategy, they are interested in browsing through greater amount of references and therefore are more tolerant of a higher percentage of irrelevance. (case 8)
The engineer from the industrial community prefers a more restrictive search strategy. He requires higher precision even when he knows that he may miss an amount of relevant information not retrieved. He tolerates this omission because his reading time is limited to a few references and he does not have the possibility to retrieve these items by a manual search in a large database.

Profiles giving search results with 90% or 100% relevance (case 1) very rarely occur. The overage precision in the system as evaluated by the users is ranging from 40% - 65%.

The increased attention to the user as significant part of an information retrieval system has stimulated various kinds of user studies.
This paper does not give an evaluation study or a user study as such. Such statistical study demands a great amount of evaluation assessments from the users and a continuous feedback analysis.
A computerized system designed to measure retrieval effectiveness in terms of a real user behaviour, gathering and maintaining data for different decision functions would be a desirable tool for a reliable study. With the increasing implementation of on-line techniques transactions files for such studies will be easier to create and maintain.

REFERENCES

1. Katter, R.V., Study of document representation. Multidimentional scaling of indexing terms. Final report. Santa Monica, California, 1967. SDC Rep. TM-3627.

2. Katter, R.V., Design and evaluation system. Annual review of information science and technology. Vol 4, 1969 p. 31-70.

3. Lancaster, F.W. & Gillespie, C.J., Design and evaluation of information systems. Annual review of information science and technology. Vol 5, 1970 p. 33-70.

4. Gluchowicz, Z., Selective Dissemination of Information from a user co-ordinator's viewpoint. Handling of nuclear information. Proceedings of the symposium on handling of nuclear information. International Atomic Energy Agency. Vienna, 1970 p. 227-237.

5. Tell, B.V. & Larsson, R. & Lindh, R., Information retrieval with the ABACUS program. An experiment in compatibility. Handling of nuclear information. Proceedings of the symposium on the handling of nuclear information. International Atomic Energy Agency. Vienna, February 1970 p. 183-200.

6. Saracevic, T., Ten years of relevance experimentation - a summary and synthesis of conclusions. American Society for Information Science. Annual meeting. Proceedings. Vol 7, 1970 p. 33-36.

7. Harmon, G., Information need transformation during inquiry: a reinterpretation of user relevance. American Society for Information Science. Annual meeting. Proceedings. Vol 7, 1970 p. 41-43.

8. Heaps, H.S., Criteria for optimum effectiveness of information retrieval systems. Information and control. Vol 18 no. 2, March 1971 p. 156-167.

CYBERNETIC ANALYSIS OF COMMUNICATION SYSTEMS

Patrick R. Penland and James G. Williams
Graduate School of Library and Information Sciences
University of Pittsburgh

SYNOPSIS

There is a pressing need for research to test the cybernetic model of communication. This need is met in part by an experimental community encounter system, where the entropy reducing components of the cybernetic model operate in a simulated social system. This system has been developed at the Graduate School of Library and Information Sciences, University of Pittsburgh for research and instructional purposes. Demographic, land use and legal data of Allegheny County (metropolitan Pittsburgh) serve as a cold-start constraints before fifteen weekly cycles carry the encounter system fifteen years into the "future." The on-line/off-line model is cybernetic in terms of its response to in-put (man or machine) and can cause new interpretations of the response by those receiving the input for the next cycle (man or machine). The encounter model of decision making involves information handling specialists in the organizational and interpersonal effects of information transfer in governmental, management and community decisions, and provides for the collection and analysis of data on a number of communications and sociological variables for research purposes.

Introduction:

Since communication is the process of creating situations (dyad, group, community) in which meanings can be engendered in people (15), then considerations of community context cannot be ignored by those in the communications profession. In general, it is the function of the disciplines to create new knowledge (largely for its own sake). On the other hand, the professions endeavor to make knowledge kinetic (information surprise) in the lives of people and call attention to lacunae in knowledge which if researched and created by the disciplines, would be socially useful (18). Without this symbiotic relationship, the disciplines become overly academic and the professions waive any leadership role for civic betterment and social accountability (19).

Until recently, no underlying theory of the societal enterprise existed which was not a static model. Even the model proposed by Hall (8), whose concept of "awareness" is particularly insightful, cannot easily be related to cybernetic theory (17). There exists however the socio-drama model of Burke-Duncan (10) which can account for the cybernetic interaction of community endeavor. The rules of the social enterprise constitute the behavioral outcomes of the roles played by adaptive control organisms as they react cybernetically to the perceptions of latent and even explicit social entropy.

The adaptive control organism has a built-in and dynamically creative antipathy to entropy. Necessity may be the mother of invention, but the imperative to escape entropy is the critical social matrix of significant symbol. Men demand structure almost more than bread as the reasoned and concluding statement of Brinton in his Ideas and Men (4) would indicate or as the pessimistic pronouncements of Toffler (20) would in a negative way try to panic the social enterprise.

Granting that men demand structure, the tragedy has been that most social theoreticians have proceeded in their analyses upon the basis of mechanistic principles (2). However an alternative position has been developed which cybernetically shapes the referents of the symbolic enterprise (9). The cybernetic structuring processing of the symbol releases participants from both the mechanisms of the mores as well as the Berkeleian solipsism. This model has been embedded in the community encounter system developed at the University of Pittsburgh with both online and off-line role components.

In the community encounter system, constraints are abstracted from societal patterns and operate within the democratic rules of gaming procedure. Participants deploy themselves according to models of communication, employing sociodrama for its function of guiding social criticism and information surprise as the motive for change. While the inclination of powerful elements in the social order is towards criticism only, information supply is positive and developmental, encouraging other participants towards a "profiles of courage" approach.

In order to realize its cybernetic potential, communication science must in Duncan's words "take into account what was communicated, the situation in which the communication occured, the kind of person who communicated (i.e. his role), the means he used, and the social purpose of his communication" (10:17). Duncan goes on to develop his thesis in the explication of several axiomatic, theoretical and methodological propositions. These are the propositions of sociodrama theory which may be refined as they are verified through research design. It is one thing to talk about research design in the sense of traditional social science (13). It is quite qnother to determine and assemble the research contexts wherein theory can be verified, let alone establish experimental vehicles for the testing of hypothesis (9).

Community Encounter System:

There is a need to know more about the communicative abilities of humans as well as the means for predicting such ability based upon experimental techniques which minimize error and maximize yield of information. The community encounter system has been developed to study an on-going simulated environment which facilitates research on the problems and component relationships of communication. The encounter system provides for a gestalt of social activity wherein the entire gamut of human communication can be studied. The encounter simulation provides a central focus in the development and evaluation of new programs and methods as well as the social justification of existing ones. It also focuses attention on the need for long range systematic research as a basis for planning situations wherein effective communication will be engendered.

For example, on an elementary level, videotapes and the computer have already proved much more effective in recording complex behavior for measurement purposes than the traditional pencil and paper methods. Of course the "cameramen" are under control of at least one researcher to ensure that its "selective eye" follows the "sense" of the dynamic encounter situation being recorded. The computer monitors the input of each role participant, provides print-outs of individual and group performance as well as statistical analyses and updates the constraints for each succeeding weekly (annual) cycle.

The encounter simulation is a unified model of communication based on a very few principles which are available in cybernetics. Cybernetics in its assumption and principles is sufficiently general to consider it a unified approach to communication. It can be a powerful tool for communication analysis. Homeostasis is not simply a response to a stimulus, but a process occuring in all parts of the encounter system and results from a transactional process among ever extending sub-systems. Cybernetic analysis can be used to diagnose interpersonal relations and categorize such transactional behavior as the communication components which accompany the self-stabilizing interactions of the entire encounter system. The transactional process requires communication of information and motives at many levels. Information processing can be related to the activities of learning, thinking and understanding.

Based upon previous knowledge as codified in the social sciences (11) and in particular information science (14) a number of assumptions underlie the research vehicle. Principally among these assumptions are the following:

> People are willing to pay for information as a social resource to the degree that it is perceived as an essential ingredient in decision making.
>
> The degree of consensus underlying the decision making process as represented on the print-out varies directly as the encounter simulation progresses.
>
> The number of people involved in decision making varies directly with the range of contacts in the communication system.
>
> People in encounter role playing situations react in a manner isomorphic with real life.

The function of criticism keeps the social system under control in terms of community objectives.

Power structure response to continuous citizen criticism follows an ordinal distribution: (a) agreement and well-wishing; (b) committees and/or commissions are appointed; (c) economic and/or legal action occurs as a result of committee or commission reports.

The advantage of cybernetic analysis is that the model transcends subject boundaries, and views people and institutions as adaptive control systems of reception, transmission, evaluation, and storage. Input is perception; judgement is decision-making; and output is behavioral action. The findings of the newer "interdisciplinary" social sciences have been synthesized in order to explicate the three components of the cybernetic model. This encounter model has been developed to replace the verbal speculation of the older social sciences with the on-line and off-line cybernetic components of reception, decision-making and behavioral patterns.

The teleological thrust of the off-line roles is toward cooperative or competitive behavior for valued ends by employing the entropy reducing functions of cybernetics for research purposes. The findings of each of the newer social sciences shed light upon, and delineate research matrices for the functions of the three components of an adaptive control organism. The following table which is somewhat more abstract than Figure 1, indicates the cybernetic components of the community encounter system which has been developed for research and instructional purposes:

Cybernetic Model Components

Perceptor	Interpretator	Effector
Organized and projected information	Role-structure (off- and on-line)	On-line Data Matrix
Mass Communication	Cognitive structure:	Land use
Organizations	organized information space	Financial constraints
Talking chains	Affective structure:	Legal constraints
	power	Demographic constraints
	government	

The encounter vehicle has been designed to explore a number of theoretical considerations. Among the propositions for which measurement devices have been initially developed are the following:

> The understanding of participants towards their roles varies directly with increased information and communicative contacts.
>
> Social power varies directly with the ability to supply information for decision making and with the number of communicative contacts.
>
> The complexity of decision making processes varies directly with the amount of information available and the number of communicative contacts.
>
> The explication of social issues and problems is a function of information.
>
> Issue awareness and problem identification varies directly with citizen involvement and monitoring of governmental decision making.

The <u>first</u> component of the cybernetic model, stimulus and reception, has been explored in the disciplines of information theory, linguistics, sign and symbol behavior which together lay a foundation for the analysis of information surprise. From a cybernetic system-model point of view there are two basic purposes for communication: to inform and to motivate. Communication functions to alter a receiver's concepts and to change his preference and feelings so that a wider range of alternatives exist for decision-making and behavioral response. This has the effect of increasing the complexity of decision making as well as extending the time interval between problem identification and behavioral action. In the construction of a message, it is necessary to build in informational and especially motivational referrents (12). However, according to the sociodrama model of communication (10) and the experience of the mass media in persuasive communication, motives are changed first and then concepts are filled in.

Human behavior is the output of a controlled, not an uncontrolled system. This viewpoint is consistent with the analytical concepts used in communication, decision-making and systems analysis which are considerations in the selector and effector stages of the cybernetic model. In addition, they are similar to the informational and motivational aspects of communication, to the opportunity and preference functions of decision-making, and to the broader cultural notions of scientific and value judgements in society. Interpersonal relationships and transactional analysis in group, organization and community endeavor are means of enlarging opportunities directly both in concept effectiveness and preference strength.

Human behavior is based on information about things, not necessarily on the nature of things. The human is able to accumulate reservoirs of information over a period of time through a process of symbolization or conceptualization. Concepts can be of things as well as of events. Once information is stored in the form of concepts, it can be tapped, or released, by cues. No action or cue can occur at a distance without some connection either of a cause-effect or a feedback relationship. The cause-effect action, no matter how complex, proceeds in one direction and is common in the natural environment. A feedback system, however, involves reciprocal relations which can be studied by the multi-correlational analysis of many variables. Correlation can not establish cause effect relations, but only show that variables co-vary. Correlation is very useful in exploring an environment to determine what variables should be brought under controlled study using other classical methods to show cause-effect.

Freedom of opportunity varies directly with an increase in number of information sources and depends on access to communication media. Mass media constitute a major stimulus source for social endeavor as well as the community encounter simulation. In order to maintain a bargaining advantage, the flow of information can be curtailed about the matters under bargaining. Where information cannot be controlled directly, perception may be modified by criticizing communication content or by raising compelling points of view against it. In any event, organized information space (library as prototype) is considered to be of surprise value as well as serving as an entropy-reducing function, whether the "beings of reason" included were developed in the past or are emerging at the present moment.

The second component, the preferential phenomena of interpretation, is explored in the disciplines of game theory, decision theory, value philosophy and psychology which taken together help to explicate the preferential patterns of the adaptive control organism. Descriptive and normative studies of decision-making have led to the analysis of aspects of human behavior in which choices are made among alternatives. For example, Simon has pursued the ramifications of decision-making and value theory in his Models of Man (6).

By way of illustrating the availability of choice among alternatives, the following changes in inputs which vary directly with outputs have been adapted from Cadwallader (9): (a) the innovation rate is a function of the legal and economic rules which organize the problem solving matrix of the social system; (b) the range of communicative contexts varies directly with the available variety of information; (c) communication capacity cannot exceed the quantity and variety of information supply; (d) the sociodrama model provides the rules (criticism) for forgetting or disrupting the highly probable organizational patterns of the power structure.

Any cybernetic system has both input and output. The effect of the environment on the system is input, while output is the effect of the system on the environment. Within the system, the effect of one component upon another is a functional relation or interaction. In a cybernetic system, the governor or selector is set by the larger or encompassing system. For example, the value system of a human individual is not calibrated by the individual alone, but mainly by the cultural system. The setting of the governor or the establishing of values for controlling behavior is a parameter of the cybernetic system determined by the larger environment and during the traditional experimental study of a system is supposed to remain unchanged. However, in the correlation analysis of nonparametric distributions, it is not essential to hold independent variables under rigorous experimental control. No apriori ideas need be placed on the distributions under study.

Almost any action of one person with another produces feedback, the action and feedback being the transactional use of power. Power is the ability to do work in the physical sciences, and the ability to satisfy wants in the social and policy sciences. The bargaining strength of one individual varies inversely with his own effective preferences and directly with the preferences of the other. Reciprocal demand is the basis of all transactions whether interpersonal or societal. Power is only as strong as the effective preference, and the limits on equilibrium within the transaction are set by the parameters of available alternative opportunities. Figures 2 and 3 indicate that a fundamental dichotomy exists between cognitive (information) and affective (power) elements of the library vs. governmental entropy-reducing complex.

Events occur and situations exist, but their import occurs in people. Problens exists only in the minds of people and become a subjective index of the fact that the world of reality did not behave the way people wanted it to do. People pay attention to the unexpected (information surprise), and the amount of attention varies directly with the degree of relationship between an individual and some event, idea or other person (s). Consequently, the power of a communication source varies directly with its ability to create situations where information surprise can be engendered.

The third component of the adaptive control organism is explored in the disciplines of general systems theory and operations research. General systems theory has grown out of the work of Lotka (1) and Von Bertalanffy (16) where generalized models, principles and laws are applied to the physical and biological behavioral systems and their subclasses. Operations research on the other hand is the applied component of general systems theory. It brings the systems approach as well as its intellectual and interdisciplinary resources to bear on organizational problems. Ashby (5), Nagel (3) and Latil (7) have explored the relation of software and hardware in operations research employing computers, automata and other control devices. In the community-encounter research vehicle, general systems theory has been taken into the future, as the encounter system investigates the impact of behavioral outcomes today upon the information processing and decision-making of tomorrow.

The Research Vehicle:

Following Ashby's postulate that cybernetics is the theory of all possible machines, organisms and combinations, the encounter system is an integrative model and a guide to the planning of individual research studies and components for an explicit galaxy of research "designs." Uniformities among the various subjects and professional disciplines have been incorporated into the simulation model. The generality of social science research findings (11) has been extrapolated to the encounter system and integrative benefits have been derived from comparing theories in apparently diverse fields.

V *Evaluation of Retrieval Effectiveness*

There is one area of communications research to which a program of investigation in the encounter simulation leads directly. Professional education is based upon an academic grading system which, while it may be pertinent to a degree program, does not necessarily ensure role definition and development. The encounter simulation creates situations in "real-life" standards where effective communicative behavior can be developed in context. In addition to evaluation which reflects content gain or progress, role definition and development can be evaluated in relation to defined group norms.

Actually from the studies already undertaken there is a reason to believe that research employing canonical correlations in the encounter simulation can be as productive as investigations carried out in highly organized experimental frameworks. "Valuc free" distributions may be logically desirable but in communications research are difficult to achieve. The advantages of research in the encounter simulation lie not so much in the degree to which the findings are rigorously verified, but rather in the large repertoire of types of studies that become possible and in the built-in stimuli to inquiry. Because of the nature of the encounter simulation, the faculty and participants have the combination of energy, insight and integrative effort that is necessary for constructive inquiry about fundamental problems (see Figure 4).

Before experimental research can be undertaken in social communications, many studies need to be conducted in contexts employing statistical controls rather than experimental controls. The distribution of variables must be identified and compared not only with one another but also with the characteristics of the components and subcomponents of the encounter system. Instead of specially selected subjects who might fit more readily into traditional experimental design, subjects are the students who register in the communications sequence. Considerable progress has been made in the development of evaluative instruments, including devices which as final instruments in one phase of evaluation can be used as initial measures at the beginning of another.

After all, there are really few indispensable elements in research design that cannot be conducted in the encounter system. The instructional personnel responsible for the library-community encounter simulation have a research point of view and possess knowledge of appropriate experimental and statistical techniques. An office of tests and measurements has readily available not only instrument to measure achievement as well as participant characteristics, but also has the resources and the means for test development, administration, scoring and statistical analyses.

Except for the randomization of treatment groups, many experimental controls can be applied after the observations are collected. Groups can be stratified or matched after the experiment is completed just as easily as before. Whenever the original groups are large enough to permit fractionalization, specifications can be imposed on the groups during the analyses. Computers can be programmed to accomplish this readily. In the matrix of intercorrelations and weightings, the unique contributions of any of the independent variables can be determined by factor analysis. Any variable can be interchanged as to whether it is to be used as a control or as a predictor. A system of experimental inquiries will generate valuable results without designing a new experiment for each new set of conditions.

Older social science methods of research have been evaluated in relation to components of the encounter system and to characteristics of senders and receivers. In addition, new methods and devices not only are being invented but they also are being evaluated in relation to "established" procedures. New methods can be evaluated relatively easily without radical disruption of the encounter simulation. For example, subject groups need not necessarily be matched. Nor do they have to be of the same size or even be concurrent. It is however necessary that the "control" groups be measured for initial level of skill or knowledge to be communicated as well as for all the characteristics under consideration as referents.

Experimental control as employed in the social sciences with human subjects uses a limited range of one or two variables. The subjects may be selected to yield predetermined distributions, or members of like pairs, or triads, or quartets may be assigned to treatment groups, usually at random. Treatment is the independent variable, residual gain the dependent. However, other variables may be used for control. Correlational analysis or canonical correlation permits the use of conventional significance tests as to whether changes in the skill or variable under study are related to method. Also through analysis of variance the practical importance of changes may be assessed to better advantage than in the usual experimental designs.

<u>Summary</u>:

Organization adds to the transactional model of communication but one element of structured cooperation which includes the development of goals and the shared experience of values or motives. With organization comes the complications of increased roles but also of relationships among the sponsors or owners (if a business or industry), the staff and the recipients. The whole society is an organization. Government is the formal organization of power. Property and legal rights constitute the set of rules about transactions. Whether the awareness of them is informal, formal or technical (4), transactions among citizens are constantly going on and continuously being formalized into "permanent" coalations. For government not to intervene in the transactions of power groups is to sanction those transactions. The scope of government is total for its society since its actions and inactions cover all possible situations.

The adaptive control organism is most productive when the cognitive and affective domain work together in a complementary manner. However such a harmonous relationship is neither always the case nor possible. Be this as it may, both realms are essential components of the community encounter system. Figures 2 and 3 delineate this cooperative vs. competitive interaction which is the matrix of the social enterprise. The encounter system is composed of participant roles and the first line of research has focused on the entropy-reducing functions of communication (information) agents. This depends not only on the manner in which information space is organized but also and especially on the way in which information is employed to cause surprise in receivers.

V Evaluation of Retrieval Effectiveness

Many specialized information processing centers are available in metropolis, but there is a serious lack of coordination. There is no one clearinghouse to which to turn. The discrepancy which exists between need and the mustering of information and of making it kinetic in the affairs of citizens is approaching a crisis. When left unattended, it is interesting that participants begin to speculate on the contributory affect of a lack of available information upon the urban disorder in the simulated encounter environment.

Information networking and library planning is an extension of the surprise value of information retrieval. Networks and interlibrary cooperation require the combination of materials, services and professional expertise which is impossible to achieve by one library alone. Information networking adds the dimension of an effective switching center at the regional, state and national levels.

Libraries in the simulated metropolis work together through regional library centers in order to avoid unnecessary duplication of resource collections and to facilitate access both to their own resources and to other libraries in the national network. Numerous other information centers exist which function to provide significant services to specialized interests. All of these benefit in varying degrees from the information transfer (bibliographic) network. Each of these may have other needs for information which primarily cannot be met from documents that stem from the usual publishing and other distribution channels.

A good amount of knowledge, particularly that related to the work-a-day life of the people, exists in records and sources that are ephemeral and not highly organized, or perhaps remain unorganized. This type of knowledge is often mission-oriented. Information is retrieved from such sources only so long as there is an obvious and continuing need, usually for a limited clientele such as a housing project or a neighborhood activist cell. There is another extensive category of knowledge which does not exist in records at all, but is available only in the minds of specialists and professional experts. This kind of knowledge becomes available only upon consultation. The information is rendered kinetic when the individual interfaces with a consulting expert over an immediate problem or specific interests such as Information Volunteers Services of the Health and Welfare Association of Allegheny County.

Finally there is the knowledge of the present moment, the near past and the impending future where the emerging needs of the people help to shape the information sources of the present moment. There are several media of communication in metropolis whose knowledge of the present, the near past and the immediate future is a constant and continuous source of information which few citizens in the area could avoid even if they wanted to do so. Most people's requirements are conditioned partially by the media and partly by their daily occupations. Consequently the information spaces sought, range widely over ephemeral and mission-oriented sources of knowledge. Only a small percentage of the information required every day is sought in the depth for which the bibliographic record is designed to serve. Of course without such inquiry from the record, however small in quantity, civilization would flounder for lack of perspective. In the long run, libraries of the record are an integral element of modern vigorous civilizations, but in the day-to-day preoccupations of people they often seem irrelevant.

Consequently a community-encounter simulated social system has been developed by the Graduate School of Library and Information Science, of the University of Pittsburgh. The encounter simulation has been built upon a cybernetic model of communication for research and instructional purposes. The power-deployment of government and the surprise value of organized information space are the central entropy-reducing components of the model being tested. Social science research methods and distribution-free statistical techniques have been used to study relations of variables in the in-put and out-put phases within and among each other as well as with characteristics of the entropy-reducing components.

References

[1] Lotka, Alfred J. Elements of Physical Biology. Williams & Wilkins, Baltimore, 1926.

[2] Parsons, Talcott. The Social System. Free Press, 1951.

[3] Nagel, Ernest. "Automatic Control." Scientific American, 187:44-7, September, 1952.

[4] Brinton, Crane. Ideas and Men. Prentice-Hall, 1953.

[5] Ashby, W. Ross. Design for a Brain. Chapman and Hall, London, 1954.

[6] Simon, Herbert A. Models of Man. Wiley, 1957.

[7] Latil, P. de. Thinking by Machine. Houghton Mifflin, 1957.

[8] Hall, Edward T. The Silent Language. Doubleday, 1959.

[9] Cadwallader, Mervyn L. "Cybernetic Analysis of Change in Complex Social Organization." American Journal of Sociology, 65:154-7, 1959.

[10] Duncan, Hugh D. Symbols in Society. Oxford University Press, 1961.

[11] Berelson, Bernard. Human Behavior, an Inventory of Scientific Findings. Harcourt, 1964.

[12] Meredith, Patrick. "Documents, Programs and Topics--Some Observations on Topic Analysis" in Focus on Information and Communication, Barbara Kyle, ed., ASLIB, 1965.

[13] Phillips, Bernard S. Social Research, Strategy and Tactics. Macmillan, 1966.

[14] American Society for Information Science. Annual Review of Information Science. ASIS, Vol. 1 - , 1966 -

[15] Dickoff, James. "Theory in a Practice Discipline." Nursing Research. 17:415-35; 545-54, (1968).

[16] Von Bertalanffy, L. General Systems Theory. Braziller, 1968.

[17] Penland, Patrick R. "Towards an Integration of Communication Theory." International Communication Association Conference, 1969.

[18] Penland, Patrick R. "Communication Producing Theory." International Communications Association Conference, 1970.

[19] Penland, Patrick R. Floating Librarians in the Community. Bookstore, University of Pittsburgh, 1970.

[20] Toffler, A. Future Shock. Random House, 1970.

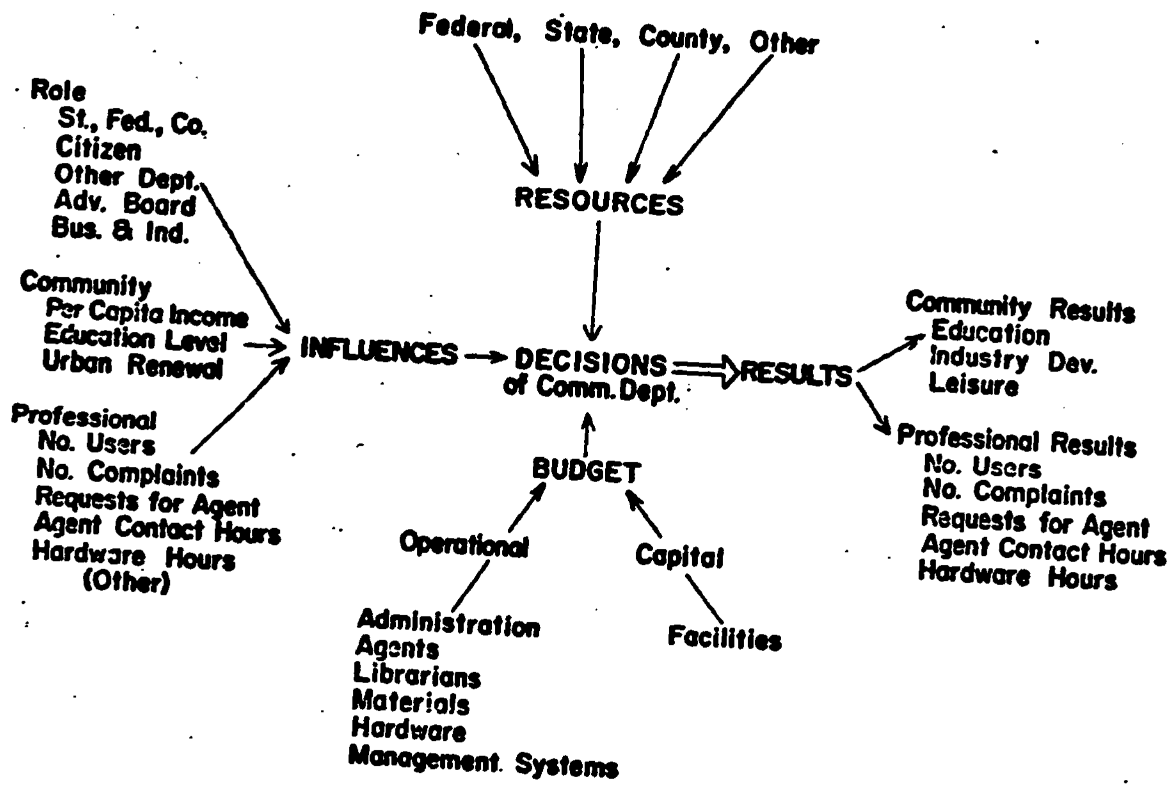

Federal, State, County, Other
RESOURCES
Role
St., Fed., Co.
Citizen
Other Dept.
Adv. Board
Bus. & Ind.
Community
Per Capita Income
Education Level
Urban Renewal
Professional
No. Users
No. Complaints
Requests for Agent
Agent Contact Hours
Hardware Hours
(Other)
INFLUENCES
DECISIONS
of Comm. Dept.
RESULTS
Community Results
Education
Industry Dev.
Leisure
Professional Results
No. Users
No. Complaints
Requests for Agent
Agent Contact Hours
Hardware Hours
BUDGET
Operational
Capital
Administration
Agents
Librarians
Materials
Hardware
Management. Systems
Facilities

Figure 1

Figure 2

Strength of Interaction

Strong

Medium

X Low but some

Direction of Interaction

CONVENTIONAL U.A. ROLES FOR COUNTY	Library Council	Comm. Budget Officer	Comm. Manager	Comm. Agent	Comm. Administrator	State Librarian	Fed. Div. of Lib. Progs.	Conventional Librarian	Business and Industry	Tax Association
State Administrator	Strong					Medium				X
Federal Administrator	Strong						Medium			X
County Council		X			X			X		Strong
County Manager		Medium			Strong			X		Medium
County Planner			Medium	X					Strong	
Urban Renewal Director			X	X					Medium	
Anti Poverty Director			X	X						
HEW Administrator			X	X	X					
Land Developers.			X							
Citizens				Medium						

COMMUNICATION ROLES

POTENTIAL CONFLICT

Conflict = Two or more groups competing for limited resources, power positions, prestige etc, resisting change or the threat of change.

High
Med
X Low

Direction of Interaction ←	1 State	2 Federal	3 State Lib.	4 Fed.Div.of Lib.	5 Manager	6 Planner	7 Communications	8 Renewal	9 Poverty (CAP)	10 HEW	11 Land Dev.	12 Citizens	13 Comm. Budget	14 " Planner	15 " Agents	16 Conv. Lib.	17 County Council	18 Council of Lib.	19 Tax Assn.	20 Business & Ind.
1 State Administrator		X	X		Med	X		X	X	X				X			X	High		
2 Federal Administrator	X			Med	Med	Med		Med	Med	Med							X	High	High	
3 State Librarian	X						X							Med		X		High		
4 Fed. Div. of Libraries		X	X				X											High		
5 County Manager	Med	Med				X	X	X	X	X	Med	Med	Med				Med			
6 County Planner	X	Med			X			Med	Med	X	High	X		X			Med			Med
7 County Communications	X	X	Med		X				X	X		X			X	X	Med	Med		
8 Urban Renewal	X	Med			X	X			X		High									
9 Anti Poverty (CAP)	X	Med			X	Med	X													
10 HEW	X	Med			X	X	X													
11 Land Developer					Med	High		High												
12 Citizens					Med			Med												
13 Communications Budget					Med				X	X						High				
14 " Planner	X		Med			X										X				
15 " Agents												X				High				
16 Conventional Librarians			X				X						High	X	High			X		
17 County Council	X	X			Med	Med	Med												High	
18 Council of Librarians	High	High	High	High			X									X				
19 Tax Association		High															High			
20 Business and Industry						Med														

Figure 3

COMMUNITY DECISION GAME

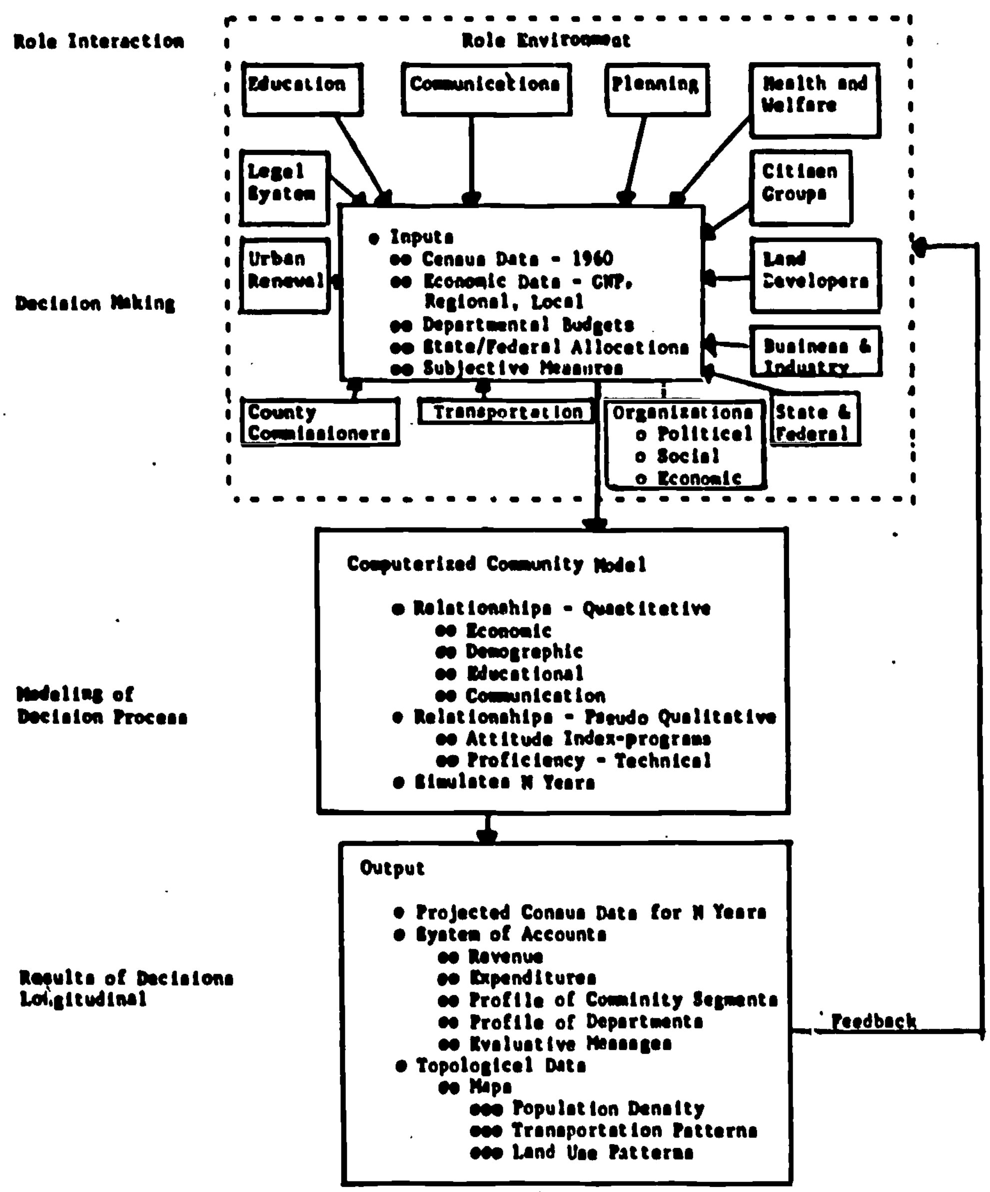


Figure 4

A GENERALIZED THEORY FOR THE EFFECTIVENESS AND UTILIZATION OF INFORMATION*

M. C. YOVITS
Department of Computer and Information Science
The Ohio State University
Columbus, Ohio 43210
U.S.A.

SYNOPSIS

A general theory is presented which defines the term "information" as "data of value in decision making". Decisions in turn result in courses of action which in turn result in actions that can be observed or measured in the physical world. A generalized theory is suggested which establishes relationships and procedures for the flow of information. Analytical relationships are suggested for describing the effectiveness of information as a function of resulting observable actions.

Introduction

One of the major problems that exists in the development of information storage and retrieval systems is that of developing methods of evaluation for these systems. Only when such methods have been established will it be possible to compare various systems as to their effectiveness and then ultimately to attempt to decide on the value to be placed on such a system. Only when effective criteria for evaluation of information systems are available will it be possible quantitatively to determine whether it is worthwhile to develop one system rather than another or even whether it is desirable to develop any system at all.

*Research sponsored in part by the National Science Foundation through Grant No. GN-534 from the Office of Science Information Service to the Computer and Information Science Research Center, The Ohio State University.

V Evaluation of Retrieval Effectiveness

Naturally in order to establish specific measures of effectiveness for information systems it is necessary to develop quantitative measures of information and to relate the information quantitatively to some measure for the utilization of information. At the present time any measure for the effectiveness of information is generally limited to the rather general, qualitative question of relevance to a user or group of users. The question of what relevance may mean or of how relevant information may be is normally not considered. The question of what value can be placed on a particular measure of relevance is not generally considered.

In the present paper a theoretical treatment which enables the establishment of quantitative measures for the effectiveness and utilization of information is presented. Some of this work has been presented in earlier papers and is accordingly summarized here. Suggestions for extending the previous work are given in order to establish relationships which enable the evaluation and comparison of information systems.

Background

The term information has many different connotations, uses, and meanings. It is sometimes used in the rather specific sense that Shannon and Weaver (1) have established. In this sense, although information and its properties and appropriate analytical relationships are well defined and quantitatively measured, the class of situations which is covered is a small part of those of interest. In the situation considered by Shannon and Weaver the context of the message is of no significance and the theory is concerned with the probability of the receipt of any particular message for various conditions of the transmission system.

On the other hand, information is frequently, perhaps almost universally, considered to be the equivalent of "knowledge". It is in this sense of the term information that information scientists and information systems designers and users are concerned, and it is this sense of the term which is the popularly understood meaning as well.

In the Shannon-Weaver sense, the term information is indeed quantifiable, measurable, and rigorously defined, but the theory is of such limited applicability as to have little utility with regard to the design of information storage and retrieval systems. In the more general sense of information being equivalent to knowledge, the term is neither rigorously defined nor measurable nor quantifiable. Furthermore, it seems unlikely that such broad and vague conceptions can yield relationships that are meaningful and useful in any real sense.

Thus, we have been concerned with the formulation and development of a theory of generalized information flow which has wide and perhaps universal applicability (2)-(5). This theory leads to a model which allows rigorous mathematical treatment. One of the motivations for the development of this model is the desire to establish analytical relationships which have as wide applicability as possible. These analytical relationships can then be used to derive principles and identify limitations on the flow and transfer of information. The model may also be used to define fundamental quantities involved in information flow, including information itself, in a quantitative, rigorous, measurable and repeatable way. In particular, the model can be used to generate specific measures to be used in evaluating the effectiveness and the utilization of information and for the comparison of information systems.

Properties and Definition of Information

To be of value it is clear that knowledge must at some time be put to use. That use is to assist in making decisions; that is, information is used in order to make decisions. Indeed, the only resource that is available to a decision maker is information, nothing more. Thus a definition of information has been proposed (2) which forms the basis for the model of a generalized information system which was mentioned above. This definition is: information is data of value in decision making. While this definition may delimit the total range of interest in information in an intellectual sense, it does have virtually universal applicability with regard to any potential applications. It is further claimed that any more general definition of the term is not amenable to the quantification and analytic conceptualization necessary to the treatment of the study of information as a science.

In the physical world, the only way in which a real parameter, say a mass or a velocity, can be measured is through an interaction of a system with its environment. If this interaction does not exist then it is impossible to measure a parameter or even to establish its existence. The same statement can be made of information. Without some type of interaction with the environment it is not possible to measure information or to know of its existence. Such an interaction will result in some kind of physical change which can be measured or observed.

A physical change or observable action will take place as a consequence of the operation of the decision-making apparatus. The decision maker will recommend courses of action which are in turn executed, yielding a set of observable actions. Since, as indicated above, the decision maker makes decisions solely on the basis of the information he receives, it is clear that this is a fundamental mechanism by means of which information is transformed into physical, observable actions.

These remarks concerning the definition, measurement, and significance of information lead to the development of a model of information flow. This model can then be used to obtain fundamental analytical relationships which can be used for measuring the effectiveness of information. It is the purpose of this paper to propose such a theoretical approach.

The Generalized Information System

Since the model of a generalized information system forms the basis for our approach, it will be useful to describe this system in some detail in order to provide the necessary background to the present research. The generalized model is depicted schematically in Figure 1.

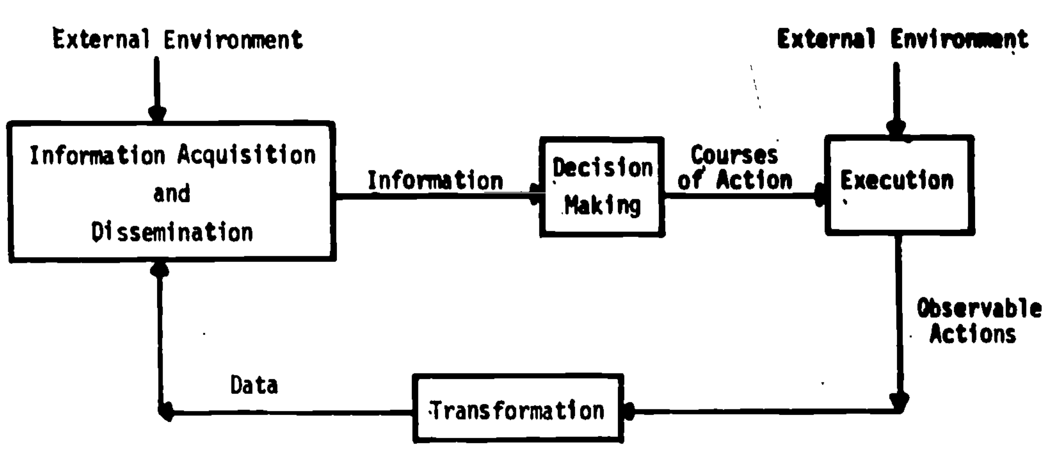


Figure 1. Generalized Information System

The system is comprised of four essential functions. These include the Information Acquisition and Dissemination function (IAD), the Decision Making function (DM), the Execution function (E), and the Transformation function (T). As has been pointed out, virtually all situations involving the flow of information can be described by this model. These situations would include the use of information by the research scientist or the development engineer, management of a large corporation, command and control of a military engagement, or such relatively straightforward and simple activities as the switching on or off of a thermostat-furnace system.

Personal decision-making problems are described by Figure 1. It is not even necessary or required that the decision-making process be a logical one. This model is applicable when decisions are completely irrational, as may frequently be the case. Each function is seen to collect, store, operate, and disseminate.

In any realizable and operational system all the indicated functions must be present, and they must be considered together in order to understand information flow or in order to establish principles, relationships, or guidelines for information flow. Just as in the analysis of any system, suboptimization or consideration of the functions independently may yield misleading or incorrect results.

In particular, the Decision Making (DM) function is a most important one and is established as the key consideration in the entire information flow process. The DM function represents any system component accepting an input from the IAD, and providing an output to E. The DM may be an individual person, an organization, a man-machine system or simply a machine system. In all of these cases, the DM transforms information into courses of action which are in turn transformed into observable actions. The input to the DM is information, some of which may be stored or held in memory. The DM makes decisions on the basis of the information available at some particular time. The decisions are made individually, serially, sequentially, or in parallel and, of course, the decision-making process may be delayed. This may be the case when more information is necessary. But the decision maker is responsible for the generation of observable actions and will eventually make decisions that will lead to these actions.

A very significant point concerns the closing of the feedback loop to the DM. In any system this loop must be closed to provide a basis for retaining or altering the courses of action disseminated and it is only on this basis of closure that the DM is able to refine or alter decisions intelligently. Such feedback is always present whether or not it is explicitly considered. Most of the presently designed so-called "information systems", however, do not consider feedback or closure to the decision maker in meaningful ways. The loop in the system is closed by the transformation of the observable actions generated by means of appropriate measuring devices, into data transmitted to the IAD which are in turn disseminated to the DM. This feedback provides the major interaction which the DM has with the total system and is the only way for the DM to probe and understand the system.

Since information is supplied <u>only</u> so that a decision maker can make decisions of some kind it is most appropriate to analyze our model by starting with that particular function.

Note that the decision maker collects three basic kinds of information from the IAD. There is information on the particular activity under consideration (that is, data which have been obtained by transforming the observable actions resulting from operations of the Execution function). There is information on the external environment over which the decision maker has no control but may have knowledge. There is also other fundamental information which the decision maker may utilize. This latter category includes reports, tables, mathematical and physical constants and appropriate relationships. The DM <u>stores</u> a data base in its memory.

The DM <u>operates</u> as follows: it develops a predictive model which it believes will transform the information received into the appropriate observable actions. This model developed by the DM may be an accurate one or it may be incorrect. It may even be irrational or illogical. Nevertheless, it is the way in which the DM believes that the only resource available it to, namely information, will be transformed to observable actions. It is only after further information is available from transformations of these observable actions into data that the DM "knows" whether its model is an accurate one or not. Secondly, the DM may alter recommended courses of action using the same predictive models, as further information regarding the particular activities involved or the resulting observable actions become available. Thirdly, the DM may develop new models as further information becomes available either concerning the observable actions or the external environment. In this situation the original model was incorrect or inadequate regardless of the courses of action suggested. The DM <u>disseminates</u> courses of action (results of decisions) which are communicated to the E function.

The E function is responsible for transforming decisions into observable actions. It <u>collects</u> courses of action from the DM and perturbations from the external environment. If it were not for these perturbations from the external environment, E would be essentially deterministic and would transform decisions in a predictable way into observable actions*. It is largely the action of the external environment which provides much inherent uncertainty in the process of transforming information into observable actions. There is also "internal noise" which introduces fluctuations into the observables. The E function <u>stores</u> nothing other than its structure or design and has no memory <u>per se</u>. It is strictly a transforming process. It operates by transforming decisions into observable actions.

Observable actions or observables are quantities which are physical in nature. As the term implies they are capable of being observed or measured. They are neither data nor information, but are <u>capable</u> of being transformed into data. This distinction is an important one conceptually. Some examples of observable actions are the heat generated by a furnace, the position of an aircraft, results of a scientific experiment, a new product developed by a manufacturing firm, or the movement of men and equipment in a military engagement. These are physical quantities, in themselves neither information nor data.

In order to become data or information, observable actions must be transformed. This is accomplished by the function which, for want of a better name, is called the Transformation function. This function is fundamentally a measuring device which transforms the physical observable actions to data. This distinction between data and observable actions, although perhaps obvious, is not a trivial one. It is an important point to keep in mind. Data are <u>transformations</u> of the observable actions.

---

*This assumes that the E function is known by the DM. This knowledge is generally obtained only by successive iterations of decisions courses of action, execution, generation of observable action, transformation to data, and then transmission to the DM, further decisions, etc.

The last function remaining to be discussed in this model is the IAD function or perhaps, more accurately, the Information Acquisition, Storage, and Dissemination function. The IAD is frequently referred to as an "information system". Although it is of course a system it is in fact an open-loop system and, as we have shown, only a part of a much more extensive closed-loop system. As has been indicated it is important to consider the total closed system as shown in Figure 1. The IAD is only a part of this system and must be treated accordingly.

In order to measure or even discuss the effectiveness and utilization of information and of information systems it is necessary to consider the total system as indicated in Figure 1. In other words the total process encompassing the decision making and execution into observables must be of concern in the determination of effectiveness and value. It is clearly impossible to determine any measure of utility without considering the total process. To consider the IAD by itself independently of the decision making, as is generally done, can not give meaningful results. Some understanding then of the class of users (decision makers) and the resulting observables must exist to develop significant measures.

The IAD function collects data from three different fundamental sources. It collects data on the particular activity under consideration (that is, data which have been obtained by transforming the observable actions produced). Data from the external environment are also collected. The external environment includes all the factors over which the decision maker has no direct control. Finally, the IAD collects basic data such as references, tables, reports, textbooks, relationships, etc. The IAD disseminates data for the use of the decision maker. These data, if used, are information within the context of our definition of information. This distinction between data and information is of fundamental importance here.

The model of a generalized information system has now been described in summary detail. All of the functions indicated must be present for any meaningful analysis of information flow. Information now can be defined and analyzed in generality by the use of this model. It is further contended that only in such a model does information really have any significance. General discussion of the storage of knowledge or data without reference to the potential uses leads nowhere, or provides minimal fruitfulness.

General Relationships Involving Information

The general definition of information which has been proposed is "Information is data of value in decision making". Data on the other hand are transformations of observable actions. They result from the measurements which have been made on physical quantities. Information is thus inherently involved in the decision making process and inherently involves transformation into observable actions of some kind.

Note that information must be a relative quantity depending upon the decision maker, the total history of the decision maker, and the particular environment involved. In this sense it is different from the physical universe where a physical quantity is absolute and independent of the particular situation. That is, a mass of 20 grams generally remains a mass of 20 grams independent of the position or momentum of the mass.

It should not be surprising that information is to be considered relative. For example, a sophisticated research paper in some technical field may have a great amount of information for a scientist who is well versed in that field but no information to, say, a housewife who would be completely unable to make use of that paper. The background of the decision maker must necessarily be important inasmuch as two decision makers given the same document may act differently, thus generating different observables.

We may represent the relationship between the information and the observables as

$$I \times T = O$$

where I is the information, transduced (T) by the DM and the E leading to the observable actions, O. More precisely, one might more properly discuss the changes in the observables

$$\Delta I \times T = \Delta O$$

Generally, I and O each would consist of a number of components and might be thought of as being represented by vectors.

Note that there will be a smallest amount of information that will cause the DM to change a decision which in turn produces a change in the observables. That is, it is the smallest amount of information which will cause the decision maker to change a course of action. As we have noted, this smallest amount of information is relative to any particular situation. This smallest amount of information is fundamental in the relationships developed. This fundamental unit we call an <u>informon</u>. It is convenient to measure information content in units of informons.

Discussion

It is just this value of the information in informons which is directly related to the significance of the information received. As we have shown, the amount of information in a message is a direct function of the change in observables caused by that information. However, the change in the observables may not by itself be of direct significance. It is the change in the observables as a function of the minimum change that would be obtained from a change in a decision that is a quantity of considerable significance. This is closely related then to the amount of information measured in units of informs.

Note that we do not specify what the various functional relationships actually may be. These must, as we have shown, be dependent on the decision making process which is involved. Thus, the relationships will differ from situation to situation and even from time to time. Or to put it in another way, information and its value, its effectiveness, its measurement, its utility, and other quantitative measures must all be relative and dependent on the particular situation which is being considered. As we have indicated previously, a particular scientific paper would contain different amounts of information for scientists with different backgrounds and it would clearly contain a different amount of information for a military commander or a statesman. Each of these individuals would make different decisions generating different observables from the same document. Of course, an individual might make no decision at all if the document is of no interest to him or if he is unable to understand it.

Once it becomes clear that information must be treated in a relative way, dependent on the particular situation which is considered, then many conceptual and specific relationships also develop. Naturally it is difficult for us to accept this type of a situation inasmuch as we are part of a physical universe where the measurement of a physical quantity is generally absolute and independent of the situation considered. Thus specific measures for the value and effectiveness of information and information systems must consider in detail the ultimate objectives to which the information is to be put and the types of decision makers who will use the information. Currently, at The Ohio State University we are pursuing research to establish more definitively some of the relationships that must exist among these various activities and parameters for various classes of users or decision makers. While it is necessary to know the decision making rules in detail in order to develop specific relationships, it is possible to develop general, useful relationships for different classes or types of decision makers. General relationships of this type are currently being sought in our research.

Conclusion

In an attempt to establish some fundamental relationships that describe the principles of information flow and transfer and which define its limitations some earlier papers have developed a generalized theory of information flow. It is believed that this theory has wide and perhaps universal applicability. This theory which has been previously described leads to a model which allows rigorous mathematical treatment. In particular, it defines information as data of value in decision making and develops mathematical techniques for measuring and quantifying rigorously information in a general sense.

V Evaluation of Retrieval Effectiveness

This paper uses that theory to suggest a procedure for measuring the effectiveness and utilization of information and information systems by relating the information to the change in observables which is attributed to the information in the document or message. It is indicated that a fundamental unit called an informon, which is the minimum amount of information required for a decision maker to change a decision, plays an important role in these determinations. If the information is measured in terms of informons, then it is closely related to the effectiveness or value of the information.

In this paper it is pointed out that further research is taking place to establish general relationships which may be used to develop criteria and quantitative measures of effectiveness for information and information systems. While detailed measures will depend specifically on the user or decision maker and the types of decisions made, it is possible to develop general relationships which are applicable to classes of decision makers.

References

1. Shannon, C.E., and W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, Urbana, 1964.

2. Yovits, M.C., and R.L. Ernst, Generalized Information Systems in A. Kent (Ed.), Electronic Handling of Information: Testing and Evaluation, Thompson Book Co., Washington, D.C., 1967, pp. 279-290.

3. Ernst, R.L. and M.C. Yovits, Information Science as an Aid to Decision Making, Computer and Information Science Research Center, The Ohio State University, Columbus, 1969.

4. Yovits, M.C., Information Science: Toward the Development of a True Scientific Discipline. American Documentation, 20 (No. 4): 369-376 (1969).

5. Yovits, M.C., and R.L. Ernst, Generalized Information Systems: Consequences for Information Transfer in H.B. Pepinsky (Ed.) People and Information Pergamon Press, New York City, 1969. Portions of this paper were presented at the National Meeting of ASIS in Columbus, Ohio, October 1968.

ON THE NOTION OF RELEVANCE IN INFORMATION RETRIEVAL*

C.V. Negoiţă
National Council for Science and Technology, Bucuresti, Romania

SYNOPSIS

The paper presents a new point of view about the "relevance" concept. The logic of information retrieval is analyzed (two-valued, many-valued and infinite-valued logic) and the truth-value is interpreted as the relevance value. The logical system which seems to be most appropriate to describe the process of information retrieval is the continuous logic suggested by McNoughton. In this case the answer of the information retrieval system is interpreted as a fuzzy set.

INTRODUCTION

Over the last twenty years, a number of attempts to furnish a theoretical foundation for the retrieval process stand out, but little accepted theory exists as yet. One of the most serious difficulties in the theory of information retrieval is the lack of adequate explications for the notion "relevance". This term is used so widely and so broadly today that it tends to connote fuzzy thinking.

Some experts pointed out that the relevance was not the subject of detailed theoretical investigation. Donald Hillman's opinion is only one of many:

"In recent theoretical discussion of the structure of information storage and retrieval systems, one major problem has received little attention. This is the problem of elucidating the concept of relevance, or connectivity. The importance of having a clear notion of this concept is hardly a matter of doubt, for retrieval outputs consisting of citations to documents which are either unrelated to each other or irrelevant to the prescription are totally unserviceable for the user. It is therefore imperative that the concept of relevance be adequately defined and a suitable place found for it in the theory of information storage and retrieval systems."[1]

This is not to say that scientists were not very much aware of the importance of relevance concept and that a fair amount of discussion was not devoted to the subject. Rather, it suggests that, historically, considerations were not directed toward and

Editor's note: This paper was retyped from a somewhat unclear original, and due to delays in the mails we were unable to get the author's corrections to the manuscript before going to press. We apologize to Mr. Negoita and our readers for mistakes or ommissions in the paper.

did not result in a theoretical, or even less, a mathematical framework for the investigation of the relevance concept*).

There is no universal agreement at present on what the primary definition of relevance should be. Some specialists go so far as to express doubt that the problem admits a definitive solution at all. Obviously, this is a topic in which no one can yet claim the final word.

With the opinion that we can get really significant insight into the structure of information retrieval systems only by applying powerful algebraic methods, this paper is intended as a contribution toward what might be a general theory of information retrieval, touching upon some features that seem to have had little attention. The research worker will find here some fresh points of departure for future research.

INDEXING

Before proceeding to a technical definition of indexing let us review briefly the intuitive ideas involved.

The retrieval process can in general be considered as a matching process between a search request and a set of documents. In so-called descriptor systems, the retrieval of items of information in response to a request involves comparing an item specification derived from the request, with the descriptions of the actual items of information. The preparation of these descriptions is normally referred to as indexing: the assignment of a set of tags or descriptors to provide an indication of the content and a means of retrieval.

An information retrieval system is then one which compares the specification of required items with the descriptions of stored items and retrieves, or lists, all the items which correspond in some defined way to that specification. The content identifiers are used during the file search as principal criteria for discriminating between retrieved and non-retrieved items. Finally, those items are identified as relevant to the request whose key information is considered to be sufficiently similar to the request keys.

Thus, a relevance assesment becomes a value judgement on a retrieved item.

We are now ready to formalize the indexing process as follows:

If d is a descriptor and x an information item, consider the proposition "the item x has the descriptor d". It is symbolized in the calculus of propositions by the single letter p. But in the calculus of functions we refer to the fact that the proposition has an inner structure and we separate the subject "item" about whom something is said, from what is said about him, namely, the property of having the descriptor d.

*Serious attempts at preliminary explications of the notion in question have been however made. In the work of Goffman[2] - substantial contribution, though rather restricted in extent, to the achievement of an adequate explication - relevance is a relationship between the search request and the set of documents, i.e. a measure of information conveyed by a document relative to a request.

As the symbol of this relation, we employ that of the mathematical function and symbolize the proposition in the form $d(x)$.[3]

The argument sign x corresponds to the subject about which we speak; the function sign d corresponds to the property holding for it. The symbol $d(x)$ indicates the inner structure of the proposition.

Since d and x are variables, the expression $d(x)$ can be given various meanings by suitable specialization of these variables. If only d is specialized, the expression $d(x)$ is not yet a proposition. The sign d is called a propositional function (predicate); the combination $d(x)$ is called a functional. If we wish to indicate the variable in a propositional function, we write $d(\hat{x})$; the circumflex distinguishes this expression from the functional $d(x)$. The expression $d(\hat{x})$ therefore, means the same as d.

The object correlate of the propositional function, the property denoted by it, is called a situational function. For instance, the property, or situational function, of redness is denoted by the predicate, or propositional function "red".

There are two ways of constructing a proposition from a propositional function, or from a functional. The first has been mentioned. We substitute a constant x_1 for the variable x, that is, we go from $d(x)$ to $d(x_1)$. This is the method of specialization. The second procedure for constructing a proposition from a propositional function is the method of quantification, which employs the all quantifier $\forall x$ and the existential quantifier $\exists x$.

The aim of an indexing system is to make propositions that inform us about the content of the documents. The proof of the truth is based on experience and observation, sources of knowledge that play no part in logic.

THE LOGIC OF INFORMATION RETRIEVAL

Two-valued logic

In a two-valued propositional calculus we write "p","q" for propositional variables, "$\sim$" for negation, "$\wedge$" for conjunction, " v " for alternation, "$\supset$" for implication, "$\equiv$" for equivalence.

Well-formed expressions built up from the preceding symbols will be called formulas.

Formulas must just have two "truth-values"; "true" and "false". They are true or false; they may not be true and false together. Hence it is intuitively clear that there are 2^n combinations of values for a formula mentioning n different propositional variables.

The given characterizations of the propositional operations are expressed in following "truth tables", in which the truth-value v(p) of a proposition p, which is either truth or falsehood, is denoted, respectively, by "1" and "0".

p	$\sim$p
1	0
0	1

$\wedge$	1	0
1	1	0
0	0	0

v	1	0
1	1	1
0	1	0

$\supset$	1	0
1	1	0
0	1	1

$\equiv$	1	0
1	1	0
0	0	1

Using these tables we may determine the value of any formula for all possible values of the variables.

V Evaluation of Retrieval Effectiveness

Let us consider the formula $p \wedge q$. Clearly we have 4 combinations of values for p and q:

for p	1 1 0 0
for q	1 0 1 0
for $p \wedge q$	1 0 0 0

If the two-valued propositional calculus is accepted as the underlying logic of an information retrieval process, the algebra ($\{p\}$, $\vee$, $\wedge$, $\sim$, u, t) of propositions, where t is the tautology and u is the contradiction, may model the storage and retrieval of information. This algebra is a Boolean algebra defined as a system (B, $\cdot$, + , $-$,0,1) for which the following axioms hold:

1.- The binary operations . and + are commutative
If $a,b \in B$, then $a.b=b.a$ and $a + b = b + a$
2.- Each binary operation is distributive over the other
If $a,b \in B$, then $a.(b + c) = a.b + a.c$
$a + (b.c) = (a + b).(a + c)$
3.- There exist in the set B two distinct identity elements 0 and 1 relative to the binary operations . and + respectively.
4.- The unary operation - is such that for every element a in set B there exists an element $\bar{a}$ in B such that
$a.\bar{a} = 1$
$a + \bar{a} = 0$

Thus, in this model*) (queries being represented by formulas and the file of documents to be interrogated by a finite set) there is a unique subset of the file which corresponds to each formula. For each member in the file the formula takes on a truth-value of "1" or "0", depending upon whether the document is or is not relevant to the query. Hence the answer consists of all members in the file which take on a value of one, i.e., all members for which the formula is assertable.

If we now consider the relationship between propositions and truth-values, in the light of the above explanation, then <u>the truth-value may be identified with the relevance value</u>. The existence of the truth-value establishes the precise nature of the connection, eliminating the confusion.

---

* An equally well-known interpretation of the same formal system is supplied by the so-called calculus of classes[4]. If a propositional function d($\hat{x}$) is given, all arguments x that satisfy d($\hat{x}$) can be incorporated in one class, the class D. Every propositional function thus defines a class. Vice versa, every class can be regarded as defined by a propositional function. The arguments x for which d(x) is true are called elements of the class D. For the expression "x is a member of D" we write $x \in D$.

Let us consider the algebra (P(X), $\cap$, $\cup$, $-$,$\emptyset$,X) of subsets $D \in P(X)$ of a set X of documents, which is also a Boolean algebra, if D is interpreted as the set of all documents indexed under term D then:

$D + E = D \cap E$	the set of all documents indexed under terms D and E
$D . E = D \cup E$	the set of all documents indexed under terms D or E
$0 = \emptyset$	the set of no documents
$1 = X$	the set of all documents

If the two-valued logic has been accepted as the underlying logic of an information retrieval process, the retrieval system may be defined formally as a relation on abstract sets Q and A representing, respectively, the set of inputs (search requests) and the set of outputs (answers).

$$S \subseteq Q \times A$$

We regard a system as a structure which receives inputs, performs some function, and produces an output. Since the inputs can be considered as causes and the outputs as effects S is a functional system

$$S : Q \rightarrow t$$

Defining the search function f of the system by the mapping

$$f : \{q\} \times X \rightarrow \{0,1\}$$

the answer A of the system is the graph of this function.*)

$$A = \{(x, f_q(x)) \mid x \in X\}$$

<u>Many-valued logic</u>

In the two-valued logic, let $\{p\}$ be the set of propositions, L_2 the set $\{0,1\}$ and v_2 a function defined by the mapping

$$v_2 : \{p\} \rightarrow L_2$$

The value $v_2(p)$ is called the two-valued truth-value of the proposition p

$$v_2(p) = \begin{cases} 1 & \text{if the proposition p is true} \\ 0 & \text{if the proposition p is false} \end{cases}$$

Such a function is a homomorphism of the algebra of propositions onto the two-element Boolean algebra of truth-values.

In a similar way, in the three-valued logic (of Lukasiewicz)

$$v_3 : \{p\} \rightarrow L_3$$

$$v_3(p) = \begin{cases} 1 & \text{if proposition p is true} \\ 1/2 & \text{if proposition p is indeterminate} \\ 0 & \text{if proposition p is false} \end{cases}$$

---

* Since $f_q(x) = 1$ if $x \in A$

$f_q(x) = 0$ if $x \notin A$

the retrieval system can be defined as

$$S \subseteq Q \times P(X)$$
$$S : Q \rightarrow P(X)$$

In other words, the retrieval system is the system (X,Q,S) where S is the retrieval function. The set of documents S(q) is said to be retrieved by the request $q \in Q$. That is, the mapping S maps each request $q \in Q$ into some subset of the document set X.

V Evaluation of Retrieval Effectiveness

In the four-valued logic (of Lukasiewicz)

$$v4: \{p\} \to L_4$$

$$v4\ (p) = \begin{cases} 1 & \text{if proposition p is true} \\ 2/3 & \text{if proposition p is possibly true} \\ 1/3 & \text{if proposition p is possibly false} \\ 0 & \text{if proposition p is fals} \end{cases}$$

In the many-valued logic (of Lukasiewicz)

$$v_n: \{p\} \to L_n$$

where $$L_n = \left\{0, \frac{1}{n-1}, \frac{2}{n-1}, \ldots, \frac{n-2}{n-1}, 1\right\}$$

Some values from L_n correspond to r levels of falsehood, other values correspond to s levels of indetermination and other values correspond to t=n-r-s levels of truth.

There is an algebra of many-valued logic that could play the role of Boolean algebra in the case of classical two-valued logic. The many-valued Lukasiewicz algebras were introduced by Moisil as models for Lukasiewicz' many-valued propositional calculus.[5]

The Moisil algebra is a lattice M which satisfies the following axioms:

1.- M is a distributive lattice with first element 0 and last element 1

2.- M has an involutive duality named negation

$$N(a \wedge b) = Na \vee Nb$$
$$N(a \vee b) = Na \wedge Nb$$
$$NNa = a$$

3.- M has n-1 chrysippien endomorphisms $a \to \sigma_i a$

a	0	$\frac{1}{n-1}$	$\frac{2}{n-1}$	. . .	$\frac{n-3}{n-1}$	$\frac{n-2}{n-1}$	1
$\sigma_1 a$	0	0	0	. . .	0	0	1
$\sigma_2 a$	0	0	0	. . .	0	1	1
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
$\sigma_{n-1} a$	0	1	1	. . .	1	1	1,

WITH $\sigma_i 0 = 0$ and $\sigma_i 1 = 1$

$$\sigma_i a \wedge N\ \sigma_i a = 0$$
$$\sigma_i a \vee N\ \sigma_i a = 1$$

4.- The elements $\sigma_i a$ form a linear ordered lattice

$$\sigma_1 a \subset \sigma_2 a \subset \ldots \subset \sigma_{n-1} a$$

5.- There is the relation

$$\sigma_i \ \sigma_j a = \sigma_j a \quad \text{for any } i,j$$

$$\text{and } \sigma_i Na = N \ \sigma_j a \quad \text{where } j = \varphi(i)$$

6.- If $\sigma_i a = \sigma_i b$ $(i=1,\ldots,n-1)$ then $a=b$

Sicoe[6] proved that a Moisil algebra is a system $(M, 1, N, \sigma_1,\ldots, \sigma_{n-1}, \wedge)$ where 1 is not postulated as last element, verifying the following axioms:

1.- $a \wedge N(Na \wedge Nb) = a$

2.- $a \wedge N(Nb \wedge Nc) = N(N(c \wedge a) \wedge N(b \wedge a))$

3.- $N(\sigma_i Na \wedge \sigma_j a) = 1 \quad , \quad j = \varphi(i)$

4.- $\sigma_i \ \sigma_j(a \wedge b) = \sigma_j a \wedge \sigma_j b \quad$, for any i,j

5.- $\sigma_i Na \wedge \sigma_k Na = N \sigma_j a \quad , \quad i \leq k \quad , \quad j = \varphi(i)$

6.- if $\sigma_i a = \sigma_i b$ $(1 \leq i \leq n-1)$, then φ $a=b$

and that a many-valued Lukasiewicz algebra with $\varphi(i) = n-1 \pmod n$ is a Kleene Algebra.

If the many-valued logic has been accepted as the logic of an information retrieval process the search function g of the retrieval system $S \subset Q \times A$ can be defined by the mapping

$$g : \{q\} \times X \rightarrow \left\{0, \frac{1}{n-1}, \frac{2}{n-1}, \ldots, \frac{n-2}{n-1}, 1\right\}$$

and the answer, by the set

$$A = \left\{ \cup x, g_q(x) \;\middle|\; x \in X \right\}$$

<u>The continuous logic</u>

In a continuous (infinite-valued) logic

$$v_c : p \rightarrow \{0,1\}$$

i.e., the proposition (or the formula) should be permitted to take on any value on $\{0,1\}$. All real numbers t such that $0 \leq t \leq 1$ shall constitute the set of truth-values which the formulas are permitted to take. This logical system, which we shall call L_c, may be constructed as follows: we shall admit a denumerable number of propositional variables denoted "p_1","p_2",...,"p_n" together with a number of propositional operators (those of the two-valued logic). Formulas in L_c are then defined by the following set of rules:

1.- every propositional variable is a formula.
2.- If F is a formula then $\sim F$ is a formula.
3.- If F and F' are formulas, then $F \wedge F'$, $F \vee F'$, $F \supset F'$, $F \equiv F'$ are formulas.

We shall denote

$$v(p) = t$$
$$v(F) = \psi$$

V Evaluation of Retrieval Effectiveness

With every formula F there is associated a truth function $\xi(t_1,\ldots,t_n)$ where $t_1,\ldots,t_n$ are the truth-values of the propositional variables $p_1,\ldots,p_n$. The value of the truth-function ξ is said to be the truth-value of the formula F. Thus, the truth-value of a formula is determined solely on the basis of the truth-values of its propositional variables.

We shall now define* the truth function of the elementary formulas by generalizing the two-valued case. Thus:

$$\begin{aligned} v(\sim p) &= 1-t \\ v(p \supset q) &= \text{Min}\ (1 - t + u\ ,\ 1) \\ v(p \vee q) &= \text{Max}\ (t,u) \\ v(p \wedge q) &= \text{Min}\ (t,u) \end{aligned}$$

The propositional operators "$\vee$" and "$\wedge$" may be defined:

$$\begin{aligned} p \vee q &\text{ as } (p \supset q) \supset q \\ p \wedge q &\text{ as } \sim(\sim p \vee \sim q) \end{aligned}$$

If formula F contains the propositional variables $p_1,\ldots,p_n$, then we shall write $F(p_1,\ldots,p_n)$ and

$$v(F(p_1,\ldots,p_n)) = \Psi(t_1,\ldots,t_n)$$

Suppose given a function $\xi(t_1,\ldots,t_n)$ defined by $0 \leq t_i \leq 1$, $i=1,\ldots,n$. McNoughton[8] has shown that for a function $\xi(t_1,\ldots,t_n)$ to correspond to a formula $F(p_1,\ldots,p_n)$, such that $\xi = \Psi$, it is necessary and sufficient that ξ satisfies the following conditions:

1.- ξ is uni-valued, continuous and
   $0 \leq \xi(t_1,\ldots,t_n) \leq 1$ if $0 \leq t_i \leq 1$, $i=1,\ldots,n$
2.- there exist a finite set of polynominals $\lambda_1,\ldots,\lambda_\mu$
   $\lambda_j = b_j + m_{1_j} t_1 + \ldots + m_{n_j} t_n$ where b and m are integers,
   such that for any $(t_1,\ldots,t_n)$, $0 \leq t_i \leq 1, i=1,\ldots,n$
   exists a j, $1 \leq j \leq \mu$ such that $\xi(t_1,\ldots,t_n) = \lambda_j(t_1,\ldots,t_n)$

If the continuous logic has been accepted as the underlying logic of an information retrieval process, in the retrieval system $S \subset Q \times A$ the search function h is derived by the mapping

$$h : \{q\} \times X \rightarrow [0,1]$$

This function maps the set X in the unit interval. Then the answer A is a set of ordered pairs.

$$A = \{(x, h_q(x)) \mid x \in X\}$$

where $h_q(x)$ is termed the grade of membership of x in A. Thus, the answer of a retrieval system is a fuzzy set, i.e., a class of objects in which there is no sharp

---

* Another continuous logical system (the probability logic L_p of Reichenbach) was suggested by Goffman[7]. In L_p:

$$\begin{aligned} v\ (\sim p) &= 1-t \\ v\ (p \supset q) &= 1-t+tu \\ v\ (p \vee q) &= t+u-tu \\ v\ (p \wedge q) &= tu \end{aligned}$$

boundary between those objects that belong to the class and those that do not.[3,9]

Let A and B be answers in a space X, with the grades of membership of x in A and B denoted by $h_A(x)$ and $h_B(x)$, respectively. Then for all x in X:

$$A = B \leftrightarrow h_A(x) = h_B(x)$$
$$A \subset B \leftrightarrow h_A(x) \leq h_B(x)$$
$$C = A \cup B \leftrightarrow h_C(x) = \mathrm{Max}\,[h_A(x), h_B(x)]$$
$$D = A \cap B \leftrightarrow h_D(x) = \mathrm{Min}\,[h_A(x), h_B(x)]$$
$$E = \bar{A} \leftrightarrow h_E(x) = 1 - h_A(x)$$

If A,B and $D = A \cap B$ are bounded convex answers then $1\text{-}\mathrm{Sup}\,h_D(x)$ is the highest degree of separation of A and B that can be achieved with a hyperplane.

The family T of answers in X is a fuzzy topology[10], which satisfies the following conditions:

$$\emptyset, X \in T$$
$$\text{if } A, B \in T \text{ then } A \cap B \in T$$
$$\text{if } A_i \in T \text{ then } \cup A_i \in T$$

The pair (X,T) is a fuzzy topological space.

REFERENCES

1. HILLMAN, D.J. — The Notion of Relevance, American Documentation, vol. 15, 26-34 (1964).
2. GOFFMAN, W. — On Relevance as a Measure, Information Storage and Retrieval, vol. 2, 201-203 (1964).
3. NEGOITA, C.V. — Sisteme de inmagazinare si regasire a informatiilor, Bucuresti, Ed. Academiei, 1970.
4. HILLMAN, D.J. — Two Models for Retrieval System Design, American Documentation, vol. 15, 217-225 (1964).
5. MOISIL, Gr.C. — Notes sur les logiques non-chrysippienes, Annales scientifique de l'Universite de Jassy, vol. 27, 86-98 (1941).
6. SICOE, C. — On Many-Valued Lukasiewicz Algebras, Proceedings of the Japan Academy, vol. 43, 725-728 (1967).
7. GOFFMAN, W. — On the Logic of Information Retrieval, Information Storage and Retrieval, vol. 2, 217-220 (1965).
8. McNOUGHTON, R. — A Theorem about Infinite-valued Sentential Logic, Journal of Symbolic Logic, vol. 16, 1-13 (1951).
9. NEGOITA, C.V. — On the Search Strategies in Automatic Information Systems, paper presented at the Sixth International Congress on Cybernetics, Namur, September 1970.
10. CHANG, C.L. — Fuzzy Topological Spaces, Journal of Mathematical Analysis and Applications, vol. 24, 182-190 (1968).

Session Five - Discussions

EVALUATION OF RETRIEVAL EFFECTIVENESS
Chairman: Dr. A. Cockx (Belgium)

MR. T.M. AITCHISON (UK): The necessity for our investigation arose when we were developing the Science Abstracts system into the integrated information service that became INSPEC. There was a need to find a subject indexing system which would be suitable for retrospective searching of the data base and SDI, as well as for the printed indexes to the abstracts journals.

We decided that there were essentially five possible index languages we might use: Title; title plus the abstract; subject heading in modified line that we use in the printed indexes in our abstract journals; uncontrolled or free indexing; thesaurus-based controlled indexing.

Obviously, it would have been ideal to be able to use the title, or the title plus abstract, since both of these were bound to be in the data base when we were producing journals. Next best would be the subject heading, as long as we were using that system.

The other languages, free-indexing and controlled language, were less atractive since these would have to be added to the data base. We had experience with both of them since we had been using a thesaurus-based controlled language in our experimental SDI service and in the SDI service in electronics before we integrated it with our main system.

On that basis, we set out on the lengthy process of carrying out an evaluation of the retrieval performance. The best one was controlled language. The next best was free indexing. Then the printed index system. Then the title and title plus abstract. In other words, the ranking was exactly the inverse of what we would have liked.

We then considered the question of the cost effectiveness and practicality. We ruled out the last three, the cheapest ones - the printed index system and the title and title plus abstract. That left us with the free indexing and the controlled language. At this stage we carried out failure analysis in recall with these two languages. The free-indexing language had a lower performance, probably due to our not having included in the question formulations all the possibilities for stating the different concepts that appeared in the questions.

The answer of course was that the thesauri we used were not based on the indexing of our documents and therefore could not be expected to provide what we required, whereas the controlled language thesaurus which had been developed was based entirely on the document file we were using. So we came to the conclusion that if, for a variety of reasons, a controlled language would be impractical, we could hope to obtain the benefits of a controlled language by using a natural-language or free-indexing thesaurus. Thus we would consider each new concept as it arose for adding to the thesaurus. But the important thing was that the free indexing would remain unchanged on our data base and this would be available for searching in natural language. This would avoid having us make any decisions in the thesaurus that would affect the data base, and would be there forever.

V Evaluation of Retrieval Effectiveness

Our evaluation was done specifically for INSPEC, in our own particular environment, and has no pretensions to general applicability. However, I wonder whether the conclusions, such as they are, may not be worth considering in relation to other subject fields.

MR. U. BLOCH (Israel): I didn't know about Mr. Aitchison's work at all until last night. In a peculiar way we have used complementary techniques in order to find out the same thing. We found in an information center which wants to use ready-made bases, usually you have bases that are either based on abstracts, or indexes, or index terms and abstracts. In this last case, it is fairly pertinent to find out whether one of the two, the indexing terms or the abstract alone, will give sufficient retrieval, or whether you need both.

Mr. Aitchison's work is based on relevance assessments for retrieved items. Now, we all know that this expression of relevance is not a clear-cut affair of either yes or no. We all know that something can be relevant if you haven't seen another paper on it, but by the time you have seen three papers, the first paper would not be relevant any more because you already know about it. For that reason we felt that we would like to develop a technique that is not dependent on assessments, but to do something that is much more objective.

We tried to minimize subjective evaluation by a two-step process. As far as the indexing terms are concerned there is really no difficulty - you have the terms. But when you want to find out what is in an abstract, you cannot simply go and say: let's compare the abstracting to the indexing terms, because then your own subjectivity will automatically say: well you can say this term is equivalent to the other. And we found very quickly that this is something highly non-objective. We decided that the extraction of terms from the abstracts has to be done in a way that will ensure that it will be consistent. Once we have developed something consistent, we no longer say it is subjective, because we find the same abstracts repeatedly give the same kinds of indexing terms.

After the first step of finding a list of possible terms, we invite a subject specialist to compare the list of terms that we have culled from the abstracts with the terms in the index. We permit him to make only three decisions: Do you find synonymous terms in the index and abstract? Do you find that one of the two has a more specific term than the other? Do you find that the term appears only in one surrogate and not in the other? This is about as far as we can go to ensure that it will be as objective as possible.

From there on, it is really a matter of counting. We assume that the greater the number of entry points the greater the effectiveness. In principle I think this right. If you go on you can of course make as many points as you want; it becomes a very noisy affair, and your precision will suffer badly. No doubt in theory this is true, but as long as you take only indexing terms from abstracts or indexing terms that have been prepared, you may assume that the terms you find in these are relevant to the subject of the item and you will not create too much noise.

Another remark I have heard is that all we have actually done by culling index terms from the abstract which we afterwards compare with the index terms prepared originally, is that we have prepared a second list of indexing terms; and as we all know, index terms are highly inconsistent. We feel that since we stressed greatly that the list culled from the abstract must be consistent between two people, this randomness has been eliminated, and we do not think that this is just another list but something more reliable. We have in fact found that after some time there is fairly good consistency between two or three indexers from the point of view of culling terms from an abstract. When we did statistical tests to see whether there is probability of chance in this, the probability was fairly low; about 1 in a hundred, and less.

MR. R. HJERPPE (Sweden): (Presentation of Gluchowicz paper). The goal of the SDI system of the Royal Institute of Technology, Stockholm, is, and I quote: "to ensure an extensive ready access to documentation in subject fields of interest for the industrial community and for the educational and research program at the universities and research institutes in Sweden." This is quite an ambitious goal, and we have tried to fulfill it.

In 1968, our system developed very slowly until we prepared our first information brochure and held our first information seminars. We have since had a steady rise in the number of inputs, during the whole period that the service was free. When we started asking for money, there were a large number of cancellations. Today we are going up again, and the number of cancelled profiles is decreasing.

In profiling, we use the usual Boolean logic. We use AND, OR and NOT operators and an operator which we might call SELECT. The SELECT operator uses a number and a letter. For example, 2B means that any 2 from B will do. Our Boolean operators operate on groups of terms instead of single terms and we have both left and right frontations. Our search terms can be of any size. In fact, in our profile we have 12 different types of terms. You can ask for all authors, or for a publisher, or for affiliation, or for the name of a journal, or for key words or classification and/or chapter heading. All in combination as you like. One of the big problems in maintaining these profiles is the interaction between the user and the documentalist. I want to stress that this interaction is something dynamic; a profile is not made up once and for all, but must be revised regularly.

Relevance judgement or precision judgement is not something static which can be given in percentages. We process totally 14 different data bases with over 1 million references per year. For those data bases which have more structure, you have to have different versions of the same profile, and ideally you should have one query profile for each base because each base is unique in content and structure. This is not practical, so we take the largest structured data bases and divide them. For example, we have a separate profile running on INSPEC. This profile first of all sub-divides INSPEC tapes into three parts: physics, electrotechnology and computer control. We then search that part which is pertinent to the question, using a normal profile and also using the classification or chapter headings which are there, in connection with natural text. This means that having first used all the natural test possibilities we broaden the scope of the query by combining the categories and the natural text terms. In this way we get more precision without losing recall, within the context of the data base.

Regarding profile construction and updating, we introduced statistical tools; for example, statistics on the number of hits for each search profile per data base run. If someone continually receives 1,000 references per run, something is wrong, because this is mostly not what the customer wants. We have some customers who are quite satisfied to be getting nothing, because this means that nobody else is engaged in same work. On the other hand, we have customers who are very eager to receive even 200, 300 references every week, although the precision might be as low as 10%; they still want it exactly this way. We have tried to explain to them that there are ways of doing it better, but they want it as it is. And we don't interfere with their wishes.

V Evaluation of Retrieval Effectiveness

We also produce a diagnosis for each output. For all profiles which give an output of more than 20 references per run, we produce a diagnosis as to which combinations of terms have resulted in hits.

We also have a thesaurus of terms which the customer uses. When queries are submitted to us, the first thing we do after having a dialogue with the customer and checking our interpretation of his query, is to select those data bases which should be used for the customer. No customer uses all the 14 data bases which are available to him. There is a table at the end of the paper which shows that of 800 profiles, 62 use only one data base, 58 use 2, 63 use 3, 201 use 4, 187 use 5, 32 use 6, and 4 use 7. Therefore, the performance for this kind of system cannot be given in a simple relevance assessment or precision assessment, because naturally the effectiveness of the different data bases will vary. The first thing which has to be done is to look at the different data bases, as in Cases 1 to 15 which are included in the paper. In these Cases, the relevance as measured by the customer varies from close to 100% to about 20%. And even when the customer receives only 20%, he may still be satisfied. So even a user evaluation in terms of precision does not actually state anything about the effectiveness of the system.

In the beginning when the service was free, the only thing the customer had to do was to give us an evaluation of the results. These results are included in the paper; out of 36,000 references, 9,000 were of immediate interest, 2,459 were of immediate interest but the customer had already read them, and 11,000 were of interest but not for immediate use. Summing up these three figures comes to something like 60% precision overall, which is quite good. It is important to stress that the interest of a customer is not static, but evolves continually. As soon as the customer receives information which is relevant to him, and which has given him something new to consider, it is quite natural that his interest changes.

Finally, the effectiveness of search from the precision point of view is very dependent on the activeness of the user. We have some users who are very active and who receive very good results, whereas we have customers who in the beginning submit a question and then do not follow it up, which makes it very hard for us to know how the system works for this particular person. We have noticed that applied research workers are more inclined to want small packages of information. We have had customers telling us: "I receive too much relevant information; I can't swallow it, I don't want this service." On the other hand, most people in basic research want a lot of output. They can afford the time to read more.

PROF. P. PENLAND (USA): I value this opportunity to talk about an approach to communication that is not often heard in the information science and retrieval field. We move with my presentation from bounded, finite systems, ISR, to finite interactive systems in a dynamic environment.

I assume that there are three components of information science. The first of these is the manual and technical operation of libraries, the usual library routines. The second is mechanized and automated ISR, information storage and retrieval. The third component which is emerging is a theoretical discipline for advancing the understanding of the information function in all the disciplines. As a result, I will not argue with those who claim that information science is a discipline. It is indeed a discipline. It is based upon theoretical considerations, and rigorous research design is used with advanced knowledge in this field.

However, despite library and ISR orientation, there are considerations which have caused us to raise questions about this communicative infrastructure called library and information science. First, it seems to me that it is difficult, if not impossible, to evaluate a system from the inside and conversely, goals for a system cannot be set within the system. It seems that no evidence has appeared yet that either the library or information processing center is not a system. In any event, I myself am often frightened at what we are doing to people in training them to be librarians and information scientists. We call it a profession, but we emphasize a service orientation, a service distribution and an information supply function, precisely those aspects which in any other area of human endeavor would be called an occupation. Of course, librarians have human relations training. But what comes out most often is public relations, salesmanship, and that kind of thing, rather than communication leadership.

The second consideration has to do with the relationship between a discipline and a profession. It seems to me that the pursuit of knowledge by means of a discipline is purposeful, perhaps even moral. It is certainly not concerned with teleology; it is not concerned with the social usefulness of the discipline. It seems to me that it makes little sense to question the validity of atomic research or computer research on the basis of the fact that these instruments may be used to destroy man. This is irrelevant to the concerns of a discipline. It is however a primary concern of a profession whether these instruments will be used for social purposes - for the betterment of man and for the good of our community.

The position that we have taken is to study the flow of information in society rather than the route of ISR, whose main concern is the creation and regeneration of knowledge in the disciplines. Users' studies, as we presently know them, are largely concerned with the scientific use of information: how do researchers use information to create new knowledge, rather than to solve social problems. The simulation that we have developed is designed to study the use of information in political and community decision-making under conditions that we feel are isomorphic with reality.

What we call library and information science is based upon the usual homomorphic and isomorphic transformation. On the other hand, the structure of the socio-economic system is based upon social power, i.e., the ability to satisfy people's preferences by means of economic profileration.

ISR is not enough to ensure that information will surprise anybody. We believe that a mediating profession is needed to create conditions which will maximize the surprise value of the retrieval of data, and data remains data until it surprises someone. Until it gives you some information that you had not expected. In other words, the important problem is to ensure that people will pay attention to the information by arousing their curiosity.

In our research we have used a variety of observational techniques. We are trying to develop new ones as we go along. Some we have used are semantic differentiation, attitude instruments, diaries, socio-metric devices both in the group and in the community, interaction analysis and the content and context analysis video tapes. We use video extensively. All group meetings are video-taped and analyzed for behavioural patterns. Mass communication is built into the model - radio, television, newspapers, etc.

I might add that we can play the simulated roles, the dynamic roles of County Commissioner, activist groups, business and industry, information specialist, floating librarians. You name it, we can play these roles both on and off-line. For research purposes, the off-line role playing is much more valuable. You can observe the behavior taking place in actual human actions.

V Evaluation of Retrieval Effectiveness

In summary, our major purposes are first, to provide a matrix within which the effectiveness of library and information storage and retrieval can be evaluated; and secondly, to identify the knowledge, attitudes and skills of a library communications profession. This is our grand purpose, our overall objective, our hope to deal with the "Gestalt" of human activity. Sounds global? It is. However, ISR is quite manageable, it is finite and it is bounded, and I am afraid most librarians and most information science specialists want a finite, bounded world. But I must point out to them that it is a highly artificial world, particularly from the view point of people who live in communities and with community problems.

QUESTION: To Prof. Wagner. Does the serial jump strategy you developed have practical applications already?

PROF. G. WAGNER (Germany): I think it is very difficult to give you general solutions to the problems of table look-up techniques. It depends on what you are doing which strategy will be best. In most cases this is done by trial and error; comparing different strategies and seeing what takes longer and what takes less time. You then take the strategy which proves to be the quickest one. Because so many parameters go into these search strategies, it is very difficult to give general solutions. What we had was one practical problem - the translation of words in strictly alphabetical order from one language to another. For this, we found this solution. But for other problems you will find other optimal solutions.

DR. COCKX: To Prof. Wagner. Do you mean that you can generalize your method to other disciplines?

PROF. WAGNER: To Dr. Cockx. It is not restricted to the medical field. It is translation of words. The SABIR system was set up in Paris with the help of our own Institute in Heidelberg. The center of the SABIR system is still in Paris and we are collaborating on the creation of an information network. The first axis of this information network is Paris - Heidelberg and other axes have since come into being such as Warsaw and Belgrade. The principle of the SABIR system, which is restricted to cancer literature, is that each country which would like to participate comes in with its own national literature in the field, puts it into the system and receives the total of the system. I think it is very economical. The central point is still Paris where most of the work is done. We in Heidelberg put the German literature into the system. But we have also been collaborating on the preparation of the system.

MRS. A. ROSENHECK (Israel): To Prof. Wagner. Who funds this project? Does every country which participates also shoulder part of the budget?

PROF. WAGNER: To Mrs. Rosenheck. We pay our costs for putting the German literature into the system. The French pay not only for the French literature at the moment, but also for the English and American literature because we have no partners yet in the US and UK. We hope to in the future and then they will pay for their own national literature. My belief is that every country should be glad to come into a large international system. It is much cheaper to pay only for the national literature than for the total of the system.

DR. COCKX: Is it preferrable to retrieve documents by searching unprocessed text or matching descriptors assigned to documents and queries? I would like to know what is the future of automatic indexing and automatic translation? Can users' needs be met better by rapid diffusion of title lists, by group profile services or by relatively slow, but high-recall/high-relevance retrieval services? And, under which conditions can the user benefit from on-line, time-sharing terminals?

DR. W.E. BATTEN (UK): To Mr. Aitchison. You used the word "best" and carefully inserted inverted commas around it. In view of what followed from Mr. Bloch it might be worth our while if you would spare just a moment to tell us what in this particular set of experiments was your criterion of "best".

MR. AITCHISON: To Dr. Batten. How do we measure "best" performance? We were talking all the time about recall, fallout and precision, so we were getting the normal Cranfield curves. We got series of curves and if one curve is more to the top right than the other, we know the recall precision all the way through is better. It is when the curves come close that one has arguments. Essentially all the performance measures were based on recall precision and fallout. But we produced curves in different ways.

MR. BLOCH: To Dr. Batten. Since we assumed that the number of terms will not be too great, we considered that the more terms the better. I stressed before that we assumed the number is not so large that it becomes noisy again. We based this on the fact that we do not extract from the full text but only from the abstract where we may assume that only relevant information has been given.

I mentioned that we assume that the greater the number of terms the greater the chance of an entry point for retrieval. This may increase the noise; however, we feel that abstracts will very seldom be noisy.

MR. B. VICKERY (UK): To Mr. Aitchison. I believe that not many of the abstracts used in the INSPEC system derive from author summaries. If this is so, is this a factor affecting the precision of the search using the abstracts? And secondly, did you collect information about the amount of overlap between the documents recalled by the different searches? I have been connected recently with one system using titles and one using descriptors to search the same base. Both performed quite well, but the overlap between the two was relatively small. They were recalling different lots of papers.

MR. AITCHISON: To Mr. Vickery. We tend in INSPEC to use author abstracts and data them as far as we can. Obviously a certain portion of the abstracts are produced separately because either there is no abstract or it is not suitable.

On overlap, we find quite a difference. In general documents either tend to come out or not to come out. If a document has a tendency to come out, it would come out with a fair number of the different languages. In other words, at each time there was this sub-set of documents that came out, you were topping up on the different languages. If you think of it, there is a reasonably good reason for that. One can imagine the subject is clear and everyone latches on to it.

MR. S. DASCAL (Israel): To Mr. Aitchison and Mr. Bloch. When you speak of titles and abstracts, I understand that you mean that no treatment has been made in the input. That means the file contains titles or abstracts that are searched in natural language without any morphological or syntactical analysis. Does this mean that you didn't use such instruments as thesauri? If in the search you used thesauri or other aids, I cannot see precisely what the different is between this and the system you have described.

V Evaluation of Retrieval Effectiveness

MR. AITCHISON: To Mr. Dascal. It is not the same to produce a controlled language or use a thesaurus. The thesaurus that provided the controlled language had been developed in setting up the data base, whereas the thesaurus that we were using for the natural language was the EJC Thesaurus and various others.

MR. W. UHLMANN (Sweden): To Mr. Aitchison and Mr. Bloch. I think one doesn't take sufficiently into consideration, considering computer costs, that a search using uncontrolled vocabulary is considerably more expensive than a direct text search.

Secondly, I think one sometimes oversimplifies the problem because, for example, if I use card string search into a character string, title for example, it is quite different from using a character string search in a string having some thousand characters such as abstract. The second needs much more character string routines. It is not sufficient that one sees that these character strings are simultaneously present in the abstract. The first might be in the beginning of the abstract and the second might be at the end, and they may have nothing to do with each other. In such cases it is essential to be able to state that these two character strings shall be in the same phrase, in the same sentence, in the same paragraph and shall be separated from each other by no more than, let's say, one word or three or four blanks and so on. This in its turn means that these character string teams become considerably more expensive. I have experience with two fairly large data bases. I can get from both roughly the same relevance figures.

A third point is that it depends very much how structured this data base is. If you have subject headings and categories and key words and abstracts, it is much easier to formulate reasonable and sensible questions, and it is much cheaper. On the other hand, if you have an unstructured data base you can get about the same results with more advanced routines if you haven't got only titles, and that depends also on what kind of title. But the unstructured data base demands approximately twice as many terms in the question as the structured data base, and that is a very considerable increase in cost. With controlled vocabulary you know what you have to ask. Let's take a question: computer traffic. On the open data base I search under the subject heading computer software, for example, only with the character string traffic. Ask it in an unstructured data base. You can imagine what you have to enumerate to get it.

MR. AITCHISON: To Mr. Uhlmann. Control versus field text in processing - I agree that one will be more efficient than the other, in computer time. We did not do context searching, but we made a somewhat abortive attempt to see whether context searching would have helped us, for instance, to overcome the fact that we got very low precision in the abstracts. In general, and that was on hindsight, the same sentence was going to help the precision, but would reduce the recall. It was surprising the number of times that what you wanted would only come out in the abstract because there was one term at the beginning and another at the end.

Regarding the structured data base, I entirely agree. Because the data base obviously only covers a certain, if multidisciplinary, field. Since we have the classification imbedded in our data base, it means that we can combine the classification with free text searching and make sure that we have partitioned the data base so that the use of the word will be very precise although it is natural language.

MR. BLOCH: To Mr. Dascal and Mr. Uhlmann. What we wanted to know is not whether this will perform well, because this obviously is a question of the kind of system you use. In order that the system can work you must have a certain amount of input data. What we wanted to know is what input data will be sufficient in order that with fairly sophisticated techniques we will be able to retrieve something. We wanted to know whether it is enough to have the indexing terms or whether it is enough to have the abstract, or whether we must use both because they do not overlap sufficiently. After we have the data base we must use a retrieval system which is sufficiently sophisticated to differentiate between character strings that are consecutive and those that are not consecutive. We were fully aware that you cannot work with pure Boolean algebra on an abstract. In 1967 we already had what we called end terms. A certain number of words have to be consecutive or must be within a given number of words or must be within a phrase or must be within a section. And only by using sufficiently sophisticated techniques can you go to an abstract. It will be disastrous to go with a simple Boolean algebra system into an abstract. There is no question that when you want to work on a long character string you must provide sufficient sophistication. Our research was based on the fact that you have indexing terms and abstracts and you want to know which is better. But we wanted to go into production and not research only and we have to live on data bases we buy from abroad. Most of them are unstructured, and therefore we have to develop our own systems. Title is very nice as long as you work from highly scientific subjects but the moment you come to engineering, the titles have no information.

<u>*QUESTION:*</u> *To Mr. Aitchison and Mr. Bloch.* Several years ago Bloomfield of Hughes Aircraft Co. did some studies on comparative retrieval effectiveness taking terms from titles and from various abstracts. He did comparative evaluation on retrieval effectiveness. I just wondered whether his work fits into this.

<u>*MR. AITCHISON:*</u> We have seen most of these studies. Our feeling was that the Cranfield approach was the most productive, although it has its critics. One of the reasons we were happy to continue in that particular way was that none of the critics seemed to have produced an alternative method which any of the other critics accepted.

<u>*MR. BLOCH:*</u> The idea was to find out whether abstracts or indexes alone are sufficient for your system through methodology which can be applied by anyone with two reasonably intelligent persons, one who knows English well and one who is a subject specialist. It is methodology only. For every data base it depends on what the indexing and abstract policy is. Only that decides whether the index or the abstract is better or whether you cannot use one without the other.

<u>*MRS. A.M. DE BUSTAMANTE (Mexico):*</u> Regarding the question of whether abstracts alone are sufficient, I would like to recall an article by R.L. Ackoff describing a very interesting experiment in which a certain number of publications selected by a special committee of the University of Pennsylvania were given to a number of students in two forms. First, abstracts of articles classified by the authors as good or poor articles were given to two groups of graduate students. The same authors then performed a test for content comprehension of the abstracts of the articles. Students were then given both the abstracts and the complete article. A further comprehension test showed that on articles of poor quality the comprehension of the students was better when they read only the abstracts; whereas on the good quality articles the students had better comprehension when they read the complete articles.

<u>*MR. W.J. NIEDERMEYR (Germany):*</u> *To Mr. Aitchison.* How large a number of items did you consider in measuring your performance and if the number was very large, how did you practically measure recall, since the procedure is very time consuming.

V Evaluation of Retrieval Effectiveness

MR. AITCHISON: To Mr. Niedermeyr. We took advantage of our particular situation. We had at the time 600 users of our SDI service. They had been receiving weekly outputs and also lists of all items in the data base. So from them we had all assessment of the relevance of some 1,200 documents. We then asked them to submit a question in the context of their SDI profile. All we had to do now was to supply them again with a copy of all the documents that they had already assessed as relevant to their profile. This gave us a relevant subset. Thus we were able to ask them on the average to look at the assessment of relevance of some 30 documents.

MR. H. FANGMEYER (Italy): To Mr. Aitchison. When you compared the abstracts and indexes, did you also foresee automatic indexing? We had performed some tests on automatic indexing in the framework of the nuclear documentation system. If we speak about indexing, we should also speak about automatic indexing, because we found that it is preferable to manual indexing. We found that the consistency between manual and computer is about the same as between manual and manual. On the basis of more than 500 documents we found that the recall and precision curve is much better for automatic indexing than for manual indexing.

MR. AITCHISON: To Mr Fangmeyer. It seems to me that the one part that the computer does not do well is to find out which of a large number of words in a paper are going to be the most relevant. I thought that this had been the experience in the automatic indexing experiments that have been carried out in the US.

MR. S. ISAACSON (Israel): I think two points have not been brought up. The first point is the indexer himself. If he is an expert in the field, the chances of relevance being high according to the retriever is going to be high. In several of the information centers I was associated with, we found that if we used an expert to do the indexing we usually got very high retrieval effects. The number of terms used varies from subject to subject, but on the average we found about 5 to 7 terms adequate. There is one other point. If one of the index terms is the category, and if you can set a weight on some of the other terms, then your relevancy of retrieval will be high. If you want to get radar measurements of the moon and the category is radar, and you set a very high retrieval margin on the word "moon", then certainly you will get very high retrieval even with a low number of terms.

MR. BLOCH: To Mr. Isaacson. We were trying to find out what happens when you use commercially available data bases. If the service from which you buy your data base employs highly specialized indexers, very possibly the indexing will be of such a high quality that it will be sufficient to use indexing terms only. We wanted to develop a methodology which could be used every time you want to incorporate a new data base in your own services. From the experience Mr. Lancaster had when he checked the MEDLARS system, it appears that it is extremely fruitful if the indexer is on the spot working with the search team because there is a kind of cross pollination, since the searcher knows what the indexer has put in and the indexer knows how the search will be done. But that can only be done if you use a data base where you do your own indexing and your own search. Many of us use systems like that, in industrial firms where you want to use your own classified material or in a large international installation like MEDLARS. Until Lancaster came along, MEDLARS kept the indexers and searchers completely separated and were surprised when it turned out that one didn't know what the other was doing.

We worried for a long time about weighted systems, but we decided on the Boolean for two reasons. It makes for a faster through-put because the moment one of the end terms of the profile is not fulfilled, you go directly to the next profile. Also it is more difficult to get a customer to make a weighted profile than a Boolean one. The scientists can feel what happens when they make a Boolean profile, even if it is a sophisticated one with strings of characters that must be together. When you ask them to put vieghts on, they feel completely lost. Recently I was teaching students on the use of computers in information work and I was surprised that even in teaching it was difficult to make them understand how the weighted system works. But it was very easy to explain how the Boolean system works.

I must mention one advantage to a weighted system. When you use a weighing system, you can update your profile automatically. So far we have not found a way to do that with the Boolean system. If someone here has some idea how to do that with Boolean, I would be extremely grateful to hear about it.

MR. C. KEREN (Israel): I would like to offer a suggestion concerning a system we are trying out just now. We take titles of articles and reports which have actually been requested by customers and extract from them the key words and compile a list of key words which we hope might later be useful. We think in fact they are useful as an aid to profiling and to making retrieval more effective. This is still very much experimental, but I think that our system might to a certain extent be an aid instead of automatic indexing.

MR. DASCAL: To Mr. Hjerppe. I'd like to know why the evaluation took so much time? About two years for each case study.

MR. HJERPPE: To Mr. Dascal. Case studies which are included were selected for three reasons. First, the customer should make an evaluation over a long period. Second, the case should concern different subjects; and third, it should concern different types of users - government, industry, etc.

MR. J. DE LACLEMANDIERE (France): To Prof. Penland. A propos de l'excellente analyse du processus de la communication faite par M. Penland, je desirerais formuler deux remarques: - Le processus de communication est presque toujours schematise par les interractions: Emetteur-Recepteur, sans toutefois que l'un se preoccupe beaucoup du contenu du message, sinon pour des raisons de cout de transmission.

Ce probleme est assez bien resolu par la theorie de l'information (information theory) qui a anticipe celle de l'information scientifique. Toutefois, en dehors de sujetions materielles, cette theorie se revele a l'experience peu feconde pour ce que nous avons a faire.

- Seconde remarque: Monsieur Penland propose une approche cybernitique de la communication, qui s'avere plausible bien d'avantage au niveau des structures, qu'a celui de la signification. Ne rappelle-t-il point que "l'homme est plus friand de structures que de pain"?

Son allusion a la "theorie des systems" stipule, que la structure represente, dans une tranche d'espace et de temps, l'equilibre en function d'un but. Cette definition nous conduit a penser que la structure est une information (LABORIT), mais il nous reste a connaitre la signification de cette structure. Et a l'heure actuelle, les systemes de traitement de l'information qui se reduisent a des systems de traitement de fichiers, pratiquement depourvus de syntaxe et de valeurs semantiques, ne nous permettent pas de nous faire une idee des significations.

V Evaluation of Retrieval Effectiveness

Ces systems sont bien eloignes des homeostats aux quel M. Penland fait allusion. La raison de cette divergence reside dans une double confusion: confusion que l'on commet entre l'informatique et l'information, et confusion que l'on entretient entre les ordinateurs et l'informatique.

Cette double confusion justife du meme coup les KWIC, les doubles KWIC, les processus de diffusion selective d'information, de diffusion automatique de publication secondaires, les reseaux documentaires non optimises, la linguistique structurale, tous produits de notre "technostructure" qui melange delicieusement necessites et aspirations sous le nom de besoins. Pour ma part je regrette que nous soyons tombes dans le piege des besoins, ce qui nous amene a effectuer un certain nombe de taches, qui ne se justifient pas, mais servent principlement a nous justifier.

On a tendence a traiter les documents comme des produits industriels mais alors que l'on connait avec precision, a la fois, la structure, le contenu, la signification et l'usage d'elements stockes en conteneurs, lors qu'il s'agit de "memoires documentaires", on ignore la plupart du temps la signification de ce qu'elles contiennent.

Cette carence est principlement due au fait que les analyseurs - programmeurs ignorent les motivations des utilisateurs et que les utilisateurs meconnaissent celles des analyseurs. Notre schema: Emetteurs-Recepteurs, s'il s'agissait d'un veritable dialogue: homme-machine, devrait comporter les feed-back d'autoregulations necessaires, dont les systemes qui nous sont proposes, sont generalement depourvus.

La communication de M. Penland est au coeur de nos preoccupations. Il en est de meme des travaux d'Osgood et d'Hayakawa qui nous ont aide en France a formuler les premisses d'une theorie psychologique de l'information.

PROF. PENLAND: I want to start off with that black box which represents ISR. It represents the sum total of recorded knowledge and unless knowledge is recorded it is not in the black box. Documents are arranged by content in a way that represents the totality of knowledge. In addition to that, the librarian-information scientist does a curious thing. He transforms that document store into a homophoric set of descriptors, backed up by abstracts in this system. And the so-called science of information says that these documents have been scientifically transformed into a set of abstracts and descriptors.

If you have transformed these documents into descriptors, why can't you reconstruct these documents from the descriptors? The answer is "Well, you know we can't do that". This, I think, says something about the bounded finite system of library and information science.

Weiner developed the adaptive control mechanism ACO. The only thing that will trigger this adaptive control mechanism is something that surprises it. If it isn't surprised, it won't pay attention. I listen to the sound of the air conditioner. I don't pay any attention to it because it doesn't surprise me. But if I were to hear music coming from the air conditioner, this would surprise me and get my attention. Until information has that surprise element you can have all the libraries and information storage units in the world, but they won't mean a thing to anyone unless there is some surpriseful relationship between the data and the individual.

Using this kind of communication model, where I, the receiver, go to a number of senders who communicate to me, and go through a mapping process and search for serendipity and if something surprises me, I pay attention. This is a model diametrically opposed to the usual ones, even the one used by Shannon, which is the kind of model where you have a sender and many receivers. Most communications theory is based upon that kind of one-way model. However, the library and information science profession is based upon another kind of model, the potential of which is not completely realized. I initiated this theory and will follow it up only as long as I am getting some data that surprises me.

MR. BLOCH: *To Prof. Penland.* I just want to remark that this idea "if it surprises me, it's information" is actually what Shannon did. But he says as long as you can predict it there is really no content in it. And it is only when something surprises you that you cannot predict, you call it information.

PROF. PENLAND: *To Mr. Bloch.* Precisely. Information centers have all the data in the world; I don't know whether they have any information at all.

SESSION SIX

selection, education and training of personnel

Authors of Papers

PATTERNS OF EDUCATION IN INFORMATION SCIENCE

Harold Borko
Professor, University of California

SYNOPSIS

Information science education has developed into a variety of forms in different countries. An analysis of these educational programs and their developing tendencies toward a more uniform set of practices suggests that a common pattern for information science education is emerging. This pattern is described and conceptualized by means of a two-part classification scheme--one dealing with the levels of competence of the trainee and the other with the orientation and content of the courses.

As one investigates the programs and procedures for education in information science, one is impressed with the diversity that exists. As one searches through the literature, problems arising from diversity are quickly encountered. Under what subject heading should one look? In the United States, the commonly used name is "information science". In the USSR and kindred countries the preferred term is "informatics", and in many European countries, the tendency is to use the term "documentation". The FID/TD Training Committee, in what we may assume to be a spirit of compromise, has published its very excellent report under the title "A Guide to the World's Training Facilities in Documentation and Information Work" [1].

In addition to the differences in the name, there is also a difference in the place where the training takes place. In some countries information science education is connected with schools for librarianship. In other countries special institutes have been organized to provide training for documentalists and other information personnel. The level of training also varies; it may take place at a secondary level, at a technical college or at a university.

The cultural, economic and political regime, particularly the difference between socialism and capitalism, influences all aspects of national life including scientific information and education. In a very enlightening article, Professor Majewski of Poland points out that [2, p. 15]:

> "In all countries with a socialist system of economy three main scientific sectors can be discerned, each possessing their own scientific and research institutes.
>
> There are scientific and research institutions existing within the framework of the
> (a) national academies of sciences
> (b) academic schools,
> (c) industry, services, organs of health protection, etc."

There are also different levels of training, and Professor Majewski goes on to explain that documentation technicians are graduates of the higher secondary schools who have, in addition, completed special documentation school courses of two years in length. Documentalists are trained in the academic schools and by means of extramural courses which provide instruction on the aims and uses of scientific information as well as in the skills of processing, classifying and disseminating documents. There are also post-graduate courses, and in the USSR one is able to obtain the degree of "doctor of information science".

The importance of the embedding educational system and how this effects schooling in documentation can be easily seen through the training policies and trends in The Federal Republic of Germany. In 1965 Professor Cremer wrote [3] that, "Because of the different education systems in the German 'lander' and the fact that the teaching of librarianship is not included in the university curricula in Germany but takes place in the so called 'Bibliothekarlehrinstitute'"...it was necessary to promote training activities at a national level. Such training became the responsibility of the Deutsche Gesellschaft fur Dokumentation in Frankfurt. This society offers special courses, examinations and a diploma.

More recently, at a meeting of the FID/TD Committee in April 1970, Mr. Marloth described some current educational developments in the Federal Republic of Germany [4]. He pointed out that now a clear distinction is made between:

(a) education and training of librarians
(b) education and training of documentalists, and
(c) education and training of information scientists.

The educational requirements for documentalists, at the highest level of competence, are a master's degree, good knowledge of the English language, one year of practice in a documentation center and a one-year part-time course at the Lehrinstitut of the Deutsche Gesellschaft fur Dokumentation. At the completion of training, the documentalist should be qualified to assume a leadership role in documentation centers.

For information science, Mr. Marloth explains that at present there exists no regulation curricula in German universities for this subject; however, a study group has been commissioned to examine the form and content of such educational training. It is anticipated that a master's degree in information science will be established and that this will be followed by a doctoral program which will be, essentially, a research degree.

Current training programs in the United Kingdom have been described in two articles, one by Farradane [5] and the other by Schur [6]. The latter paper is particularly noteworthy for it mentions some of the problems in information science education that are faced in the U. K. and, as the author says, no doubt in other countries as well. Included among these problems are the following:

(a) It is difficult to interest a sufficient number of good young scientists for training as information specialists;
(b) Existing first degree, and even post-graduate programs, are of a general nature lacking an adequate balance of specialized courses in, for example, mechanized information retrieval, reprography, management, etc.;
(c) The expansion of courses in information science is limited by the shortage of qualified teachers;
(d) Research and development in scientific and technical information takes place outside the schools of librarianship and information science, and there is a problem of how to integrate the results of this work quickly and efficiently into classroom teaching.

Efforts are being made to alleviate these problems, but the solution is still a long way off.

It is tempting to continue in this vein and to discuss the separate educational programs of other countries such as Japan [7], the USSR [8], Canada [9, 10], the United States [11, 12, 13, 14], Israel [15], and the countries of South America [16] and Africa. It is tempting, but the cataloging of differences will not be particularly enlightening nor can a brief mention do justice to these individual developments. One can easily become impressed with the differences that are present and loose sight of the similarities that are equally prevalent and which are, perhaps, of greater importance.

VI *Selection, Education and Training of Personnel*

The diversity and experimentation in educational procedures that exist are part of a necessary stage of development, but now as international activities expand there is a need to increase the similarities among the training programs and decrease the differences. As educators from different countries exchange experiences, and as they become more familiar with each others' achievements, there will be a mutual influencing of the training programs, of courses, text books and teaching aids. Information is an international commodity; no one country has a monopoly. Information must be disseminated across boundaries, and international cooperation is essential. The desire for a free exchange of information contributes toward cooperative efforts, and these in turn bring about a greater uniformity in the training programs and in the competence of the information specialist.

Efforts for furthering cooperation and for raising educational standards are facilitated by such organizations as FID, UNESCO, ICSU/AB, and others. Not only do these organizations provide a forum for the exchange of ideas, but they engage in projects that have an international scope. A project of particular significance for information scientists, and one which will have a great effect on the standardization of educational training is UNISIST. The UNISIST program envisions [17, preface pp. VI and VII]:

> "The unimpeded exchange of published or publishable scientific information and data among scientists of all countries; The cooperative development and maintenance of technical standards in order to facilitate the interchange of scientific information and data among systems; The promotion of cooperative agreements between and among systems in different countries and in different areas of the sciences for the purpose of sharing workloads, and of providing needed services and products; and the development of human and information resources in all countries as necessary foundations for the utilization of machine systems."

Among the recommendations that the UNISIST study committee made is one [17, pp. 56-57: recommendation for #13], which deals with manpower development.

> "For all nations to take an active share in the operation of international information systems, a concerted effort is needed to provide information specialists, librarians and documentalists, with improved educational facilities... Attention should be given to the desirability and feasability of internationally oriented training and educational assistance programmes, which might include proposals for pooling resources, where needed, in a number of regional education centers."

What then are the common trends that can now be identified as one studies the educational programs of information science in various countries of the world? Can these programs be catagorized and classified? The questions are difficult ones to answer because the programs are still evolving and any attempt at classification will necessarily have to ignore significant individual differences. However, all programs are not unique; similarities do exist and more will emerge. Perhaps the best way of escaping the dilemma is to avoid classifying the existing education practices and to develop a conceptual taxonomy which will help us to understand the full range of programs in information science education that seem to be developing. With this in mind, I would like to suggest a two-part classification scheme; one dealing with the levels of competence of the trainee and the other with the orientation and content of the courses.

Education in the information sciences, documentation or informatics will be seeking to train people at three different levels. These are:

(a) the Information Technician: Requiring a baccalaureate or first degree
(b) the Information Specialist: Requiring a master's degree
(c) the Information Scientist: Requiring a Ph.D. or equivalent doctorate degree.

Education at the technician level will be of a very practical nature emphasizing the analysis, organization and dissemination of information by both manual and automated techniques. People with this degree will make up the majority of the workers in documentation centers.

The information specialist is equivalent to the professional documentalist. He will receive more intensive basic training plus additional training in languages and will acquire competence in a subject specialty which may be one of the sciences or management. Graduates of this program will be qualified to take a leading role in the organization and management of the documentation center.

The doctorate degree in information science is a research degree, and the information scientist will be qualified to conduct research and to teach.

Four different orientations to the organization of an information science curriculum can be identified [13]. These are:

(a) the library oriented curriculum;
(b) the computer science oriented curriculum;
(c) the systems oriented curriculum; and
(d) the information science oriented curriculum.

The library oriented curriculum tends to identify information science with documentation, and these programs are frequently located in library schools. The emphasis is on the use of computers in libraries, and the course of study includes library automation, circulation control, on-line cataloging, etc. Also included are courses in information retrieval and dissemination, automatic indexing and abstracting, and the automation of various library functions.

The computer science oriented curriculum places primary emphasis on the computer. Such programs are generally located in engineering schools, and the course of study stresses the mathematics and logic involved in computer hardware and software design. Applications are also a topic of study, but these too emphasize engineering and mathematics, as in the use of computers in simulation studies or to represent recursive functions. Other applications that might be included are linguistics, information retrieval, data base organization, public health, etc.

The systems oriented curriculum emphasizes the methodology of systems analysis as it applies to one or more institutional environments. The curriculum deals with the analysis and design of information systems, information retrieval, library networks, etc. Stress is placed on the management and decision-making aspects of information systems rather than on the operational and service needs.

The information science oriented curriculum is based upon the premise that information science is a distinct discipline with its own methodology and research interests. Schools having this orientation emphasize theoretical courses drawn from the formal disciplines of mathematics, logic, and perhaps linguistics. Courses in computer utilization, language processing and data base management systems are also included. The emphasis of the entire program is theory, methodology, and research.

These are the discernible trends in information science education. The course structure may have any one of four different orientations, and within any given structure, students may receive training at three different levels of competence.

The challenges facing education in the information sciences of this particular moment in history are very great. The universities throughout the world are in a furment. Students everywhere are clamoring for more relevance in their courses of study. However, information science is in an enviable position. What can be more socially acceptable than studies that lead to improved methods for gathering, processing and disseminating information to all who need it and for the betterment of all mankind?

The study of information science is now in its infancy. The pattern for its future development is reasonably clear. It is up to us, the educators in this field, to nuture and accelerate the growth of information science into a fully accepted discipline with its own body of theoretical knowledge and useful application. This is the challenge that we face and that we must meet.

References

1. Ostrowska, K. and Leach, A., A Guide to the World's Training Facilities in Documentation and Information Work. First Edition, 1965, The Hague, FID, 218 pp. Second edition, 1969, 294 pp.

2. Majewski, Z., Trends in the training of documentalists in Countries of Eastern Europe. In Proceedings of the 65th Congress of the International Federation of Documentation, Washington, D. C., October 1965, p. 13-15.

3. Cramer, M., Integrated training policy for documentalists on a national level: new trends in the Federal Republic of Germany. In Proceedings of the 65th Congress of the International Federation of Documentation, Washington, D. C., October 1965, p. 61-62.

4. Marloth, H., Education and training of information specialists in the Federal Republic of Germany; a paper presented at the FID/TD Committee meeting, Frankfurt/M, April 1970, 7 p.

5. Farradane, J., Progress in documentation. Journal of Documentation, Vol. 6, No. 3, September 1970, p. 261-265.

6. Schur, H., Education and training of information specialists in the United Kingdom; a paper presented at the FID/TD Committee meeting, Frankfurt/M, April 1970, 5 p.

7. School of Library and Information Science, Keio University, Catalogue, 1970-1972.

8. Mikhaylov, A. I.; Chernyy, A. I.; and Gilyarevskiy, R. S., The development of Informatics in the USSR, Nauchno-Tekhnicheskaya Informatsiya (Scientific and Technical Information) Series 2, No. 11, 1967, pp. 3-18.

9. Scientific and Technical Information in Canada, Special Study No. 8, Ottawa, Canada, 1969.

10. Piternick, G. and Brearley, A., Canadian training of scientific information workers. In Proceedings of the International Conference on Education for Scientific Information Work, London, April 1967, pp. 73-86.

11. Belzer, J. (editor) Information Science Education: curriculum development and evaluation. American Documentation, Vol. 20, No. 4, October 1969, pp. 325-376.

12. Rush, J. E. (editor) Education and Information Science--a symposium. Journal of the American Association for Information Science, Vol. 21, No. 4, July-August 1970, pp. 265-302.

13. Borko, H. and Hayes, R. M., Education for information science (documentation). In The Social Impact of Information Retrieval: Seventh Annual National Colloquium on Information Retrieval, Philadelphia, Penn., May 1970, Medical Documentation Service, The College of Physicians of Philadelphia.

14. Borko, H., A study of the needs for research in library and information science education. Institute of Library Research, University of California, Los Angeles, October 1970, p. 291.

15. Schur, H., Post-graduate science information course in Israel. Unesco Bull. Libr., Vol. XXIV, No. 5, September-October 1970, pp. 251-265.

16. Sanz, M. T., The training of librarians and documentalists in Latin America. Unesco Bull. Libr., Vol. XXI, No. 6, November-December 1967, pp. 318-326.

17. Unesco and ICSU/AB. UNISIST synopsis of the feasibility study on a world science information system. Unesco, Paris, 1971.

TRAINING IN LIBRARIANSHIP AND
INFORMATION SCIENCE IN ISRAEL

Lydia Vilentchuk
Head, Dept. of Training and Publication
National Center of Scientific and Technological Information, Tel Aviv

SYNOPSIS

Development of education in all trends of librarianship and information science on non-graduate and graduate levels in Israel is reviewed, and statistical data about graduates for the last 14 years is given. Forecasts of needs in trained man-power for the next 10 years are made. Importance of education of users of information from an early age is stressed.

To given an indication of the number of potential users of the services of special librarians and information scientists, as well as the required trends and depths of their training, let us first look at some pertinent Israeli statistics.[1,2,3]

Total population in 1971 (approx. 17% non-Jews)	-	approx. 3,000,000
Primary schools	-	1,442 (456,356 pupils)
High schools	-	246 (138,850 pupils)
Institutes of higher learning	-	7 (37,343 students)
University-trained manpower in science and technology, including medicine and agriculture	-	approx. 25,000
Master's degrees granted in 1969 in science and technology, including agriculture	-	400
Doctor's degrees granted in 1969 in science and technology, including medicine and agriculture	-	220
School libraries	-	exact figures not available
Public libraries	-	611
Special libraries	-	283
Academic libraries	-	7

VI Selection, Education and Training of Personnel

These figures provide the statistical background to the subject. However, there is a great gap between the number of potential and actual users of libraries and information services. The Israeli public lacks habits and training in the use of libraries and recorded information. This can be explained by the curricula of the primary and high schools of the country, which do not provide systematic training in the use of libraries and thus do not foster information-mindedness from an early age. The great influx of new immigrants with their language difficulties and different backgrounds in schooling and reading habits is also not conducive to the creation of information-mindedness among the population.

Most secondary schools, primarily administered by the local authorities, have recently established libraries, but most of these libraries are not in the charge of trained librarians and no formal orientation classes in the use of libraries and of published material are offered to the pupils. The situation is even worse in primary schools, which are administered by the Ministry of Education and Culture. In most cases these schools do not have any libraries at all, as the Ministry does not earmark special budgets either for library premises or for the salary of a librarian, and only small budgets for book purchases. The collections are in the charge of a teacher-librarian who devotes not more than 8-12 hours per week to this work. It is interesting to note that the Ministry considers the work of a teacher-librarian less qualified than that of a teacher, and equates one hour of teaching assignment to one and a half hours of a teacher-librarian[4].

It is curious that although Israel is among the most book-minded nations it does not yet have a Public Libraries Law, and the provision of library services to the general public still depends on the goodwill and insight of local government bodies, trade unions or other institutions, which undertake to provide such services. However, the Ministry of Education and Culture has recently begun to establish public library networks in various areas, especially those settled by new immigrants, and has prepared a draft law on public libraries to be submitted to the Knesset (Parliament).

Until 1956 the only body providing education and training of semi-professional manpower for public libraries was the Libraries Division of the General Federation of Labour (Histadrut), whose activities in this field started long before the State of Israel came into being. The first courses offered by this Division were one-day courses, which later developed into two weeks intensive boarding courses. Their syllabus included simple cataloging, Dewey classification, physical care of books (very important in a hot climate), bibliography of reference and standard books, with emphasis on Judaica and modern Hebrew literature.

An important step forward in the training of non-graduate librarians has been made by the Israeli Library Association, founded in 1953. In 1964 the Association introduced professional examinations in librarianship and is now the only body which grants recognised certificates of proficiency in the library profession on a non-graduate level. The Association is now engaged in a revision and enlargement of its examination syllabus. Examinations take place at two levels: "A" general basic knowledge in library techniques, "B" a more advanced level with specialization in either public librarianship or special librarianship or special librarianship (see Appendix 1). Successful candidates are granted a certificate of "Assistant Librarian" or "Librarian" respectively. Between 1965 and 1970 the Association granted 170 certificates of "Librarian" in the public libraries, and 30 in the special libraries[5].

Courses leading to these examinations are offered by various organizations: the Library Association, municipalities, and the Libraries Divisions of the Histadruth and of the Ministry of Education and Culture. For the "B" level examination in special librarianship the only body offering preparatory courses is the Extramural Studies Division of the Technion - Israel Institute of Technology, which in 1958 started evening courses for this purpose in Tel Aviv and Haifa. These courses became a regular feature; at first of one year's duration, they were later extended to two and a half years and now comprise over 300 instruction hours.

In the training of librarians at the graduate level, an important development occured in 1955. Through the good offices of UNESCO, Prof. Leon Carnovsky of the University of Chicago Library School was invited to prepare a report on the library situation in Israel[6]. On his recommendation a Graduate Library School was established at the Hebrew University in Jerusalem in 1956. The duration of the course is one academic year and the minimum entrance requirement is a Bachelor's degree. At present a diploma course, the management intends to extend the studies to two years, leading to a Master's degree in librarianship. This development is expected to take place in the near future. In the second year of study the curriculum will accomodate different trends in librarianship; at present the School does not cater adequately to the training needs of special librarians and certainly not of information scientists.

Out of a total of 300 graduates during the 14 years of the School's existence, only 15% are employed in special libraries, as against 65% in academic libraries[7]. This can be explained partly by the curriculum of the School, which emphasizes conventional library techniques, sources of information in the liberal arts and Judaica and the administration of academic and public libraries (see Appendix 2). Partly it is due to the educational background of the students, as the majority (68.7%) possessed first degrees in the liberal arts and in Judaica, 23.3% in the social sciences, and only 9% in the natural sciences. No student had a degree in engineering or medicine.

To fill the gap, the Center of Scientific and Technological Information of the National Council for Research and Development, in collaboration with the Weizmann Institute of Science in 1968 began to offer a postgraduate course in information science. In 1968 and 1970/71 (in 1969/70 no course was offered) 7 and 13 students respectively graduated (13 is an estimate, as final examination results are not yet available). 45% of the students possessed degrees in the natural sciences, 35% in engineering, 10% in social sciences and 10% in the liberal arts.

This course is financed by the Ministry of Labour and the Ministry of Defence. The graduates are in great demand and more positions than students are available. All students found employment either with research institutes or with industry. However, because of budgetary problems and the difficulty in recruiting teachers, the course will probably be continued on a biennial basis only.

As schools and institutes of higher learning do not introduce their students to the use of libraries and literature, all training in librarianship and information work in Israel has to devote a disproportionally large part of the curriculum to basic information on the use of catalogues, classification schemes, reference books, etc., - skills the students should be familiar with before embarking on professional studies.

The newly trained information scientist will only seldom be able to rely on the support of an experienced librarian or have the benefit of starting his career under the supervision of a senior. A considerable amount of his time will be devoted to the supervision of routine techniques. This, and the development of oral and written communication skills - sadly neglected at all levels of education in Israel - have influenced the shaping of the syllabus. (see Appendix 3).

Some institutes of higher learning have recently begun to introduce library orientation periods for undergraduates and it is hoped that in the near future such periods will become a general feature at all of them. The Center of Scientific and Technological Information last year sponsored two short courses in the use of chemical literature, one at an undergraduate level at the Hebrew University and the other at a graduate level at the Weizmann Institute of Science (see Appendix 4).

The need for trained information workers on all levels in the near future can be evaluated from the following:

The libraries Division of the Ministry of Education and Culture conservatively estimates that after the Public Libraries Law becomes effective, a total of 150 new librarians (both graduate and non-graduate) will be required to staff new public libraries during the first five years, and approximately the same number for extension of services of the existing libraries and for replacements.

How many professional librarians will be required to staff the primary and secondary libraries is difficult to predict. This depends largely on the readiness of the Ministry of Education and Culture to train pupils at an early age in the use of information media. But even lacking the official blessing of the Ministry of Education it can be safely predicted that the number of school libraries and librarians employed in them will grow in the very near future. Presently only 2% of the Hebrew University Library School graduates are employed in school libraries[7].

Considering the need for information scientists, and starting from a cautious evaluation that for every 100 potential users - scientists, medical men, engineers and senior management personnel - one information scientist will be required during the next 10 years, 250 information scientists will have to be trained during this period. In this estimate the number of 25,000 potential users is taken as static; the yearly growth of the user community by approximately 650 equals the number of those retiring from active professional life. Any addition of users through immigration is presumed to be covered by an additional information scientist from the same source. A ten year period is chosen because the integration of information scientists into the economy of the country takes time, as the Israeli scientific and technological community is not yet information-minded and therefore not prepared to absorb a large number of information scientists today. As to special librarians (see statistics p. 1) - these are presently very few (see statistics p. 2 and 3), and they are to be found mainly with research institutes and university and government departments. In industry only the largest organizations posess special libraries, and very few have information centers. But there is currently a change of attitude, and plans are underway, partly initiated by the industry itself, to establish regional and/or subject or mission oriented information centers requiring trained information workers.

This leads to the conclusion that there is room in Israel for at least one additional graduate school of librarianship and/or information science. The first, oriented to academic and public libraries (including school libraries), should train students with a liberal arts background; The other, oriented to special librarianship and information science, should accept students with a science background, and their respective curricula should be planned accordingly.

Conclusions

From the experience gained during the 15 years of the existence of a formal training network in Israel (a country conventionally regarded as somewhere between a developed and a developing country), the following conclusions can be drawn:

a) Beginning at an early age, possibly from the primary school level, pupils should be instructed in the use of libraries and information media.

b) Extensive training in library and information science techniques should be provided on a non-graduate level - creating a cadre of good technicians to support the qualified reference librarian and information scientist, and thus free them from routine library work.

c) On a graduate level, different schools should be established for training of personnel for work in academic, public and school libraries, and for those for work in special libraries and information centers. To try to embrace in one syllabus both trends seems inadvisable as they require entirely different educational backgrounds and different personal qualifications.

References

1. Israel Government Year Book 1970/71. Jerusalem, Government Press, 1970
2. Central Bureau of Statistics. Statistical Abstracts of Israel, 1970. Jerusalem, Government Press, 1970
3. Directory of Special Libraries in Israel, 3rd ed. Tel-Aviv, Center of Scientific and Technological Information, 1969. (Guides to Sources of Information Series no. 1)
4. G. Bergson, Member of Pedagogic Secretariat for Primary Education and Teacher's Training - Ministry of Education and Culture. Personal communication from: Internal regulations
5. Israel Library Association, Examination Committee. Internal Statistics 1971
6. Leon Carnovsky Report on a Programme of Library Education in Israel. Paris, UNESCO, (1955). (Technical Assistance Special Reports, no. 1)
7. J. Rothschild, Director of the Graduate Library School, Hebrew Univ., Jerusalem, personal communication

<u>Appendix 1</u>

EXAMINATION SYLLABUS OF THE ISRAEL LIBRARY ASSOCIATION
FOR THE ATTAINMENT OF A CERTIFICATE AS
"ASSISTANT LIBRARIAN" AND "LIBRARIAN"

A. *"ASSISTANT LIBRARIAN"* operative from 1972.
The candidate has to be over 18 years of age; possess an Israeli matriculation* certificate or its equivalent, and minimum skill in typing in Hebrew or English.

The examination is conducted in writing only and consists of four papers of 3 hours each.

I. Cataloguing.
(i) kinds of catalogues; (ii) definition of terms in Hebrew and English; (iii) filing rules; (iv) preparation of catalogue cards based on ALA rules - determination of entry and heading, descriptive and analytical cataloguing, serials and periodicals.

II. Classification.
By Dewey Hebrew Abridged Edition, which includes special extension on Judaism and Israel.

III. Basic Reference Tools for the Librarian.
(i) part of the book: title page, contents page, introduction, main part, illustrations, notes, bibliographies, indexes, binding, cover; (ii) various groups of reference books: general dictionaries, bibliographies (international, national, trade, etc.), encyclopedias, geographical sources of information, who's whos, periodicals; (iii) criteria for evaluation of reference books; (iv) techniques of reference work; (v) a list of 58 reference works the candidate should be familiar with.

IV. Administration.
(i) general knowledge about different kinds of libraries: national, academic, special, public, children's; (ii) administrative, technical and reader's services; (iii) premises and equipment; (iv) the profession of librarianship in Israel; (v) main international and national organizations concerned with librarianship.

B. *"LIBRARIAN"* in the Special Libraries trend.
The current syllabus which has been operative since 1965 will remain in force till 1972 inclusive. The new syllabus, operative from 1973, is still under deliberation, so that only a general outline can ge given.

The candidate has to possess a certificate of "Assistant Librarian" and prove that he has worked at a recognized library for at least one year and that his knowledge of English is not less than required by the British General Certificate of Education O level.

---

* Matriculation corresponds to the British General Certificate of education - A level.

The examination is conducted in writing only and consists of 5 papers.

I. Cataloguing.
Special problems in cataloguing.

II. Classification.
By UDC - Hebrew Abridged Edition, and Medium English Edition (when published). Some knowledge about LC, Bibliographic and Colon Classification schemes.

III. Bibliography.
General bibliography on an advanced level, natural sciences and technology (incl. medicine and agriculture), social sciences and economics. Covers classical textbooks, abstracting services, subject dictionaries, encyclopedias, who's whos, international and national organizations and their publications.

IV. Management of Special Libraries and Information Centers: staffing, budgeting, reporting, internal publications, reprography (systems, equipment and operation).

V. Documentation.
Principles of indexing and abstracting; literature search techniques and preparation of bibliographies; principles and operation of manually operated feature and edge-notched cards; use of KWIC and KWOC indexes; use of Citation Index; use of thesauri; basic knowledge about computer applications to library work; handling of collections of special material: reports, patents, standards, trade catalogues, etc.

Appendix 2

SYLLABUS OF THE GRADUATE LIBRARY SCHOOL, HEBREW UNIVERSITY OF JERUSALEM

The syllabus comprises 610 instruction (contact) hours and is divided into required and elective courses of 430 and 180 hours respectively.

Required courses	
1. Cataloguing	70 hrs.
2. Library classification (general and Judaistic schemes)	60 hrs.
3. Hebrew literature	20 hrs.
4. Bibliography (general, Judaistic, and humanities <u>or</u> social sciences <u>or</u> advanced Judaistic <u>or</u> natural sciences)	120 hrs.
5. History of libraries and librarianship	20 hrs.
6. Administration and management	100 hrs.
7. Introduction to data processing	40 hrs.
	430 hrs.

Elective courses	
1. Advanced cataloguing	40 hrs.
2. Advanced classification (subject headings and Judaistic subjects)	40 hrs.
3. Bibliography (children's literature or oriental studies)	10 hrs.
4. History of books, writing and printing	60 hrs.
5. Book and paper preservation	10 hrs.
6. Comparative librarianship	20 hrs.
7. Public libraries (incl. adult education, readers studies)	70 hrs.
8. Academic libraries	20 hrs.
9. Children's libraries	30 hrs.
10. Archives	20 hrs.
11. Preparation for printing	20 hrs.
12. Reprography	20 hrs.
	360 hrs.
the student must elect	180 hrs.

Students with no practical library experience must carry out one month's work at a recognized library, and after passing all the oral and written examinations are granted a diploma of "Qualified Librarian".

<u>Appendix 3</u>

SYLLABUS OF THE GRADUATE COURSE IN INFORMATION SCIENCE
OF THE CENTER OF SCIENTIFIC AND TECHNOLOGICAL INFORMATION
in cooperation with the
WEIZMANN INSTITUTE OF SCIENCE

The syllabus comprises 560 instruction (contact) hours, divided into 4 main subject areas, and a 6 weeks "Special Study", carried out individually by each student under the supervision of a tutor selected by the programme director of the course. Home assignments are calculated on the basis of 2 hrs. for each contact hour.

1. Sources of Information 143 hrs.
(i) General reference sources and basic search techniques - 33 hrs.; (ii) General sources of information in science and technology: books, periodicals, abstracting services, reports, patent specifications, standards, conference proceedings, trade catalogues, learned and professional societies - 40 hrs.; (iii) Subject-oriented sources of information: life sciences, earth sciences, exact sciences, engineering and technology - 70 hrs.

2. Information Processing 152 hrs.
(i) Cataloguing: types of catalogues, filing principles, union catalogues, card services - 22 hrs.; (ii) Library classification: Dewey, LC, Bibliographic Classification, and a detailed study of UDC including index construction - 34 hrs.; (iii) Advanced classification and indexing: theory of classification, analysis of retrieval systems, types of indexes - chain, hierarchial, etc., KWIC and KWOC indexes, citation indexes, feature and edge-notched cards, descriptor languages, file organization and coding, chemical codes, search procedures - 44 hrs.; (iv) Oral and written communication; report and review writing, abstracting, transliteration, preparation of bibliographies, proof-reading, editing, oral presentation (in Hebrew snd English) - 52 hrs.

3. Use of Machines in Information Processing 145 hrs.
(i) Reprography: contact and reproduction methods, equipment, materials, processes and costs - 18 hrs.; (ii) computer fundamentals: computer hardware, system analysis flowcharting, computer languages (emphasis on COBOL) - 75 hrs.; (iii) Computer applications to library and information work in libraries for household purposes, construction of indexes, IR systems in use: MEDLARS, DDC, EURATOM, AEC, NASA, MARC CODATA, SDI, etc., costing - 52 hrs.

4. Information Management 104 hrs.
(i) National and international information and library networks, institutions, cooperation - 8 hrs.; (ii) Introduction to statistics and survey techniques - 36 hrs.; (iii) Organization and methods - 14 hrs.; (iv) Management of special libraries and information centers - 48 hrs.

Miscellaneous 14 hrs.
Introduction to course; cybernetics; technological forecasting; visiting lecturers from various organizations; visits.

Students passing all written examinations and submitting a satisfactory "Special Study are granted a diploma.

Appendix 4

SYLLABUS OF THE COURSE IN THE USE OF CHEMICAL INFORMATION SOURCES GIVEN TO GRADUATE STUDENTS AT THE WEIZMANN INSTITUTE OF SCIENCE, 1970

1. Scientific Information: General definition and concepts of information, value of scientific information; methods of exchange of information and their efficiency; scientific literature as a source of information.

2. Chemical literature: (i) Sources: encyclopedias, monographs, data collections, compendia, journals, dissertations, patents, industrial and government research reports, etc. (ii) Needs of the user: background information, current awareness, retrospective information.

3. Tools: (i) Abstracting Services. (ii) Indexes: author, subject, classified, formula, patent, citation, keyword; their advantages and disadvantages, and their various uses. (iii) Chemical Nomenclature. (iv) Chemical Notation: purposes and uses, fragmentation codes, computer readable nomenclature (Dyson, Wiswesser, Hayward systems), connectivity tables.

4. Mechanized Information Systems: information store, search and retrieval tools; description of some operational mechanical information systems; the C.A.S. registry system: mode of operation and uses; commercially available tape services for search and retrieval, their mode of operation, and efficiency.

5. Information Centers: Development of regional information centers, their tasks and services, construction and operation of small specialized information centers: purpose and use. Industrial information centers: personal files, modes of handling and operation.

EDUCATION AND TRAINING OF LIBRARY PERSONNEL: THE INDIAN PROGRAMME

Prof. P.N. Kaula
University Librarian & Head of the
Department of Library Science,
Banaras Hindu University, Varanasi

SYNOPSIS

Traces briefly the history of education and training of library personnel in India which started with a training class in 1911, developing into a postgraduate Diploma course in 1937, and a Master's Degree programme in 1948. This was followed by research facilities and special courses on documentation. Describes the five levels of courses to produce semi-professionals, professionals, specialists, and scientists.

1 HISTORICAL BACKGROUND

1.1 Library Schools: Old Phase

Library education in India started with the opening of a training class in the earstwhile State of Baroda by W.A. Borden, an American Librarian and a student of Melvil Dewey in 1911. He had come to India to be the Director of the State Library Department in 1910. In 1915, another American Librarian Asa Don Dickinson who had been invited to organise the Punjab University Library, Lahore (now in Pakistan), started a library school at the University Library. Both the training programmes in Library Science by the two American Librarians and students of Melvil Dewey had a tremendous impact on library education in the country in the initial stages. This resulted in the introduction of Dewey Decimal Classification, Anglo-American Cataloguing Rules, and other American practices. The Punjab School continued till the Partition of India in 1947 and one of its earliest products who became the Librarian of the Imperial Library now called National Library at Calcutta, conducted a Diploma course from 1935-44 on similar lines.

1.2 Library Schools: New Phase

A new phase in library education started from Madras. This was due to Dr. S.R. Ranganathan, the then Librarian, Madras University and since 1965, National Research Professor in Library Science, who started the first systematic programme in Library Science education under the auspices of the Madras Library

Association in 1929. This course was taken over by the University of Madras in 1931 which was converted into a full-time postgraduate Diploma course in 1937. The theory and practice evolved in India was taught on comparative basis with other well known practices in this school. The impact of this school in the country has been very significant.

Another university that provided for Library Science training was Andhra University which started a course in 1935. Banaras Hindu University established its Library School in 1941 and started the postgraduate Diploma course from 1942. This was followed by Bombay School in 1944, Calcutta School in 1946, Delhi School in 1947. After the Independence of the country, a number of library schools have been established in various States in India. Today there are library schools administered directly by universities or through their affiliated colleges in 13 States. The States of Bihar, Jammu and Kashmir, Himachal Pradesh, Orissa, and Nagaland have no university library school of their own. An account of library schools had been given by the author in 1962[1] and 1967[2]. Carl Hintz had also made his own evaluation in 1963[3]. A chronological table of the library schools in universities in India is given in Appendix 1. The State-wise distribution of library schools is given in Appendix 2.

2 LEVELS OF THE COURSES

The levels of Library Science courses in India are as under:-

S.No.	Course	Award	Level
1.	Certificate in Library Science	Cert Lib Sc	Semi-Professional
2.	Bachelor of Library Science/ Diploma in Library Science	B Lib Sc/ Dip Lib Sc	Professional
3.	Master of Library Science	M Lib Sc	Specialist
4.	Special Courses	ADRTC/ADR	Specialist
5.	Research in Library Science	PhD	Research

2.1 Certificate in Library Science Courses

These courses are mainly conducted by professional associations to train semi-professionals who hold lower positions in libraries or join as school librarians. Some of the universities and colleges have been conducting this course of 3 to 4 month's duration. A list of library schools at this level is given in Appendix 3.

2.2 B Lib Sc Courses

B Lib Sc course is conducted by universities to produce professionals. This course was earlier named postgraduate Dip Lib Sc (Diploma in Library Science). The change was brought about as a result of the recommendation of the Review Committee of the University Grants Commission[4]. This course is of one year's duration and is open for graduates only.

2.3 M Lib Sc Courses

M Lib Sc course was first instituted in the country in 1948 at the University of Delhi. It is a two-year course - the first year being a first professional course which was then a Dip Lib Sc and now B Lib Sc. Banaras Hindu University started this course in 1965. This was followed by two other universities at Bombay (1967) and Chandigarh (1970).

2.4 Special Courses

The Documentation Research and Training Centre, Bangalore instituted a course in Documentation of 16 month's duration in 1962. It provides for post-course apprenticeship as well. The course has been recognised equivalent to M Lib Sc. The Indian National Scientific Documentation Centre, New Delhi, and the Indian Association of Special Libraries and Information Centres, Calcutta, are offering Diploma Courses in Documentation and Reprography, and Special Librarianship respectively.

2.5 Other Courses

The Women's Polytechnics set up in the country at various places are conducting Diploma in Library Science (Dip Lib Sc) courses of one to two years duration. These courses are open to undergraduates as well. At present the Polytechnics at Ambala, Bangalore, Chandigarh, Delhi, Jullundur and Rourkela are conducting these courses.

2.6 Research

Facilities for research in Library Science, leading to PhD are available in the University of Delhi, and DRTC, Bangalore. The minimum duration for earning a research degree is two years.

3 METHOD OF ADMISSION

3.1 B Lib Sc Course

The minimum educational qualifications prescribed by all universities for admission to B Lib Sc is a basic bachelor's degree. However, in several library schools, the demand of candidates seeking admission is so heavy that graduates do not get admission in such schools. In that category are the schools at Delhi University and Banaraa Hindu University. In Banaras Hindu University, the applicants seeking admission for 1969-70 and 1970-71 have been 519 and 480 respectively. It, therefore, conducts an Admission Test to select candidates. In case of working librarians and deputed candidatea, higher academic degrees are not insisted. In several schools a specific percentage of seats is fixed for working librarians. Madras and Mysore schools insist on a pre-training apprenticeship of 4 months. Some schools do not attract talents and have also a large number of seats for admission. Some figures in this connection are as

under:
These relate to 1969-70 and is based on a questionnaire.

1.	Total number of students enrolled in 20 universities	1,000
2.	Highest enrollment in a Library School (Calcutta)	103
3.	Lowest enrollment in a Library School (Lucknow)	15
4.	Percentage of enrollment per school	35
5.	Percentage of Women students	53
6.	Average age of students	26
7.	Admission by academic attainments:	
	B.A.	53.4%
	B.Sc.	15.4%
	M.A.	30.00%
	M.Sc.	1.2%
8.	Knowledge of Foreign Language	3.00%

3.2 M Lib Sc Course

The minimum qualifications prescribed for admission to M Lib Sc course is the basic professional degree or postgraduate diploma (B Lib Sc/Dip Lib Sc). A minimum of 50% marks in the basic professional degree is required at Banaras and Delhi Schools followed by an Admission Test. Delhi, however, does not screen its own students. Bombay School does not prescribe any minimum marks for admission but stipulates a minimum of three years experience in an approved library.

DRTC course prescribes an intensive screening through a Test for a very limited number of seats.

4 SCHEME OF PAPERS

4.1 B Lib Sc

The curriculum for B Lib Sc course is more or less in conformity with the scheme of papers recommended by the Review Committee of the University Grants Commission[5]. The core papers are the following:

1. Library Organization
2. Library Administration
3. Physical Bibliography and Book Selection

4. Documentation, Bibliography and Reference Service

5. Library Classification (Theory)

6. Library Cataloguing (Theory)

7. Library Classification (Practice)

8. Library Cataloguing (Practice)

9. Records of Practical Work

Some schools offer additional subjects as given below:

1. Paper on General Knowledge by Bombay, Karnataka, Poona, Shivaji and SNDT Schools.

2. Paper on Survey of Books and Ideas by Osmania School.

3. Paper on Library Development in India by Andhra School.

4. A Project Report with 100 marks by Punjab School.

5. An elective to be selected from (a) Children's libraries; (b) High School, College and University Libraries; (c) Public Libraries; and (d) Special Libraries at Osmania School.

Provision of additional papers may be justified but the duration of the course of one year does not provide sufficient time to go through the professional content of the courses. Moreover, General Knowledge is implied in Reference Service if it is covered properly.

4.2 <u>M Lib Sc</u>

The Syllabus of M Lib Sc course in Banaras and Delhi Schools is on the pattern suggested by the Review Committee of the University Grants Commission[6]. The subjects prescribed are:

1. Universe of Knowledge: Its Structure and Development.

2. Depth Classification (Theory)

3. Advanced Library Cataloguing (Theory)

4. Depth Classification (Practice)

5. Advanced Library Cataloguing (Practice)

6. Any one of the following:

   (a) Public Library System
   (b) Academic Library System
   (c) Research and Technical Library System
   (d) Documentation

7. Literature Survey in any one of the following:

   (a) Humanities
   (b) Natural Sciences
   (c) Social Sciences

8. Current Problems in Librarianship as one of the following projects:

   (a) Field Studies
   (b) Survey Projects
   (c) Documentation Project
   (d) Literature Survey

Bombay School, however, does not follow the UGC pattern. The scheme of papers of this course is as under:-

Group A: Written Papers.

   (a) Comparative Studies in Librarianship (two papers)
   (b) Research methods and Documentation Techniques

Group B: Observation

Group C: Oral and Viva Voce.

Punjab School has adopted a modified pattern of written and practical papers. The scheme of papers of this course is as under:

1. Evolution and pattern of knowledge

2. Advanced Reference Service

3. Documentation and Information Retrieval

4. Advanced Classification and Cataloguing (Theory)

5. Advanced Classification and Cataloguing (Practical)

6. Bibliography and Classics

7. Library Organization and Administration

8. Dissertation

5 SYSTEM OF EVALUATION

The examination for both the degrees are held at the end of the course. There is a written examination as well as practical. Banaras School has introduced the system of having examiners for practical papers like other disciplines in a university. Rajasthan School follows this practice for B Lib Sc course. Delhi and Chandigarh schools have introduced the semester system. Periodical class tests and internal assessments are also made. Terminal examinations are held for evaluating the student's performance during the course of study. In B Lib Sc course, professional tour of various libraries in the country is made for observational study and evaluation of this report is added to the marks of the final examination. Some schools have a regular programme of tutorials, seminars and observational study.

The gradation of the results is in the category of Class I, Class II and Class III All the schools prescribe a minimum of 60% marks to obtain a Class I, but there is variation in the percentage of marks to obtain a Class II which is generally a minimum of 50%. In Banaras, Burdwan, Calcutta and Gujarat Schools, Class III has been abolished. The minimum for a Class II in Burdwan School is 45%, Calcutta School 40% and in Gujarat School 45%. The minimum marks for a Class III varies between 30% and 35% in some schools.

In M Lib Sc, the gradation is between Class I and Class II. Papers 7 and 8 prescribed in Banaras and Delhi Schools are in the form of dissertations which are evaluated like other dissertations in a university. Distinction in a subject is awarded if the marks obtained are above 75%.

6 THE FACULTY

The library schools in India started with part-time teaching staff who were basically appointed librarians to work for their respective libraries. The status that teaching could give to a librarian, induced a number of them to start Library Science courses managed basically with part-time teachers drawn from the library. Aligarh School has been the first school to appoint full-time Lecturers in 1957. Of course, Dr. S.R. Ranganathan had been the full-time teacher at the Delhi School from 1947-55. The Review Committee on Library Science of the University Grants Commission (UGC) recommended full time teaching staff in library schools. The author of this paper was appointed the first full-time Reader (Assistant Professor) at the Delhi School in 1958. The pattern of teaching staff varies from one school to another. While some schools have appointed full-time faculty members, some are still managed entirely by part-time teachers who are basically practising librarians. A large number of schools have both part-time and full-time teachers. In some schools librarians from other libraries and persons from other disciplines are associated with the teaching as visiting teachers.

The position of the library schools is also governed by the provision of teaching staff. Wherever there are part-time teachers, the concerned library school is managed by the librarian of that university. 16 university schools and 4 polytechnic schools are headed by full-time teachers. In rest of the schools, librarians are also heads of the teaching departments. The schools with full-time Professors as heads are at Banaras and Madras. DRTC has Dr. Ranganathan as Hon Professor with one more Professor as its head. An analysis of the teachers is given in Appendix 4.

The analysis in the above appendix shows that there are 80 full-time teachers including full-time heads of departments. The total number of part-time teachers is 88.

6.1 Library Facilities

Library facilities in most of the schools are available to teachers and the students. The Delhi School has a book collection of about 6,000 volumes and receives 112 periodicals on Library Science. It has a professional Librarian to manage it. Banaras School has a collection of 5,000 volumes and it is subscribing to 58 periodicals on Library Science. At the lower level, there are newly established Library schools which have practically no library on Library Science of their own. Even in matters of text books and the number of copies of basic books, they depend on the assistance of neighbouring libraries. Rajasthan School has built up a Workshop for the use of students.

7 EVALUATION OF THE COURSES

7.1 Number of Schools

The number of library schools offering B Lib Sc is fairly adequate. Out of 83 universities functioning in India, 34 universities have departments of Library Science. Four universities offering M Lib Sc is not, however, adequate. With the growth in the number of professionals, India will need 164,263 M Lib Sc and B Lib Sc holders at the end of Sixth Five Year Plan. The estimates of manpower requirements have been worked out in the following seven categories by a panel at DRTC, Bangalore and given in a table forming Appendix 5.

S.No.	Category (Class)	Professional required	Minimum Qualifications
1.	Class A	6	PhD or M Lib Sc & Master's degree in some other subject
2.	Class B	7	"
3.	Class C	55	"
4.	Class D	656	M Lib Sc or B Lib Sc and a Master's degree in some other subject
5.	Class E	3,593	"
6.	Class F	159,946	B Lib Sc (Second Class)
7.	Class G	79,689	Certificate in Library Science

The annual output from library schools is about 1,000 for B Lib Sc and about 50 for M Lib Sc. With the accelerated Library development programmes in the country, the product from schools have to be increased.

7.2 Contents of the Courses

Library schools generally offer a one set of B Lib Sc course which is endorsed by the University Grants Commission. As is evident, the scheme of papers do not take into consideration the developments taking place elsewhere and also within the country. They do not meet the new functions of the libraries adequately. The emphasis in the scheme of papers should be in developing adequate and up to date library services in institutions. In M Lib Sc, there should be more elective papers so that students turn out specialists in one area of Library Science or other. The needs of special libraries, university libraries and documentation centres requiring specialised services should be covered at this level. Special courses on school librarianship and special librarianship should be introduced. M Doc (Master in Documentation) courses be instituted along with M Lib Sc.

7.3 Standardization of Courses

There should be an accrediting agency for degree courses. The Indian Library Association being inactive, the initiative has to be taken by the Indian Association of Teachers of Library Science (IATLIS) established in December 1969. It held its annual seminar on Teaching Methods in Library Science in December 1970. The growth of Library Schools should be checked and an evaluation of the teaching and other facilities in existing schools made so that higher teaching standards are maintained. The Association (IATLIS) is conducting a survey of library schools and plans to bring out a "Library Science Year Book" from 1972. Similarly an accrediting agency is required for Cert Lib Sc courses. No evaluation of these courses has been made so far. The Review Committee of UGC was concerned with the evaluation of professional courses conducted by the universities. It has, however, left the teaching at the semi-professional level to library associations who by themselves are unable to reach a particular standard. While there is a certificate course of one year's duration conducted by the Delhi Library Association from 1955, there are similar courses of 3 to 4 months duration. Papers offered also range from three to eight in these courses. Some colleges have also started such courses, though no recognition has been accorded to them by the universities with which the colleges are affiliated.

There is a brain drain in library science as well. Several librarians who qualify even with M Lib Sc degree migrate to the United States to join MLS programme once again. There is therefore need to examine the equivalence of M Lib Sc with MLS degrees offered by American and Canadian universities. M Lib Sc in India is a two-year full-time professional degree programme while MLS in U.S.A. is not so. Perhaps a Committee of Association of American Library Schools and the Indian Association of Teachers of Library Science can resolve this issue. Recently British Library Association has taken up the issue of equivalence of FLA and MLS with the Association of Canadian Library Schools.

7.4 Duration of Courses

The duration of a year to both the professional graduate degree courses and postgraduate degree course is not sufficient. In the M Lib Sc course, especially the two projects covering intensive knowledge-content and research make it difficult for students to read extensively for specialization within the limited period. The students do not find sufficient time to do practical, experimental and project work; and at the same time, study for the number of papers prescribed in one year. The duration of both the degree courses be increased to three years instead of two years as at present. The project work may be completed within the next academic year after the evaluation of theory and practical papers. Dr. Ranganathan has agreed on this issue[7]. For admission to M Lib Sc, a minimum three years experience in library work should be prescribed so that the students are familiar with the application of theoretical knowledge learnt during B Lib Sc period.

8 RESEARCH FACILITIES

The Library schools even at the advanced level, have little facilities for doing research. The Delhi school has not attracted sufficient candidates for its doctoral programme. DRTC is the only school in the country where research in various branches of Library Science and Documentation is being conducted by its faculty members and is able to attract candidates for doctoral programme. There are very few teachers in other library schools who engage themselves in research and publish papers or books. Some teachers in the Banaras, Delhi, Gauhati, Punjab, and Rajasthan schools have produced works on library science. Facilities to do research for development of the subject by teachers in library schools should be adequately provided by the university authorities and the University Grants Commission. Sufficient fellowships and scholarships should be made available both to the students and teachers in library schools. The seminar on the Teaching of Library Science in 1966, recommended that library schools actively take to research on library problems, faced by librarians in their day to day operations[8].

Jesse H. Shera strongly pleads for research programme in library schools. He says "Librarians know very well <u>how</u> to do what they do, but they never concern themselves to any great extent with <u>why</u> they do it."[9] Library education in India must be research-oriented at all levels.

9 TEACHING METHODS

Library schools conduct training programme mainly through lecture method. In some schools, other methods of teaching like seminars, tutorials, colloquia, project work, and observational study are emphasized. All the schools should do minimum of lecturing and impart instruction through other effective methods of teaching. The teachers of Library Science should also undergo a training programme on teaching methods[10]. Teachers should have worked in some good library before they become teachers. They should engage themselves constantly with the practical problems in a library and demonstrate them to the students. The University Library should be used as a laboratory of library science by teachers and students. Teaching should be integrated with theory and practice of Library science.

10 RECOMMENDATIONS AND ATTAINMENT

Library education in India has been a recurring theme of debate and discussion among the members of the profession. Four All-India seminars have been held on the subject during the last decade. Two of these seminars were held in Delhi (1960 and 1966)[11], one at Bombay (1965) and one at Banaras (1966)[12]. Library education has also been discussed in other professional conferences. A survey of the Delhi School has been made by Morris A. Gelfand[13]. A survey of library education has been made at Banaras School[14]. Karl Hintz and the author have made evaluation of the training programmes in 1962, 1963 and 1967. Review Committee of UGC (with Dr. Ranganathan as Chairman and the author as one of the expert members) set up in 1961, has made a number of recommendations. Another Committee for M Lib Sc courses has been appointed by UGC last year. The recommendations of the seminars and of UGC Committees have gone a long way in developing library education in India on proper lines.

India has the unique distinction of starting M Lib Sc Programme in 1948 when no other university in the entire Commonwealth had provision for it. This position of India continued till 1965. India has also established an endowed Chair in Library Science at the Madras School in 1958 which is perhaps the only endowed professorship in the East. India has also appointed Dr. S.R. Ranganathan as 'National Research Professor in Library Science', thus recognising library science as a discipline at par with other academic disciplines in the country. This recognition according to Mr. P.N. Kirpal, Chairman, Executive Board of Unesco, is "unique in the world"[15]. By forming the Indian Association of Teachers of Library Science, India has achieved another distinction in Library Science by instituting a Fellowship for study in library education at DRTC, Bangalore. This has been made possible due to the donation received from a distinguished teacher from the United States. It is proposed to have exchange of teachers with other countries to give new shape to library education.

BIBLIOGRAPHICAL REFERENCES

1. KAULA, P.N. Education for Librarianship in India (Library World, vol. 63, 1962, pp. 223-228).
2. KAULA, P.N. An Evaluation of Education for Librarianship in India (Unesco Bulletin for Libraries, vol. 21, 1967, pp. 182-183).
3. HINTZ, Carl Education for Librarianship in India. Occasional Papers, University of Illinois. 1963.
4. UNIVERSITY GRANTS COMMISSION (New Delhi). Library Science in Indian Universities: Report of the Review Committee on Library Science. 1965, pp. 36.
5. Ibid., pp. 17.
6. Ibid., pp. 20-21.
7. RANGANATHAN, S.R. and KAULA, P.N. Some Points on Library Education in India (Herald of Library Science, vol. 5, 1966, pp. 302-303).

8. KAULA, P.N. Teaching of Library Science: Delhi Seminar (Herald of Library Science, vol. 5, 1966, pp. 295).

9. SHERA, Jesse H. Librarians against Machines (Wilson library bulletin, 1967, pp. 69).

10. UNIVERSITY GRANTS COMMISSION. (New Delhi): Op cit.

11. KAULA, P.N. Teaching of Library Science: Delhi Seminar. Op cit. pp. 278-297.

12. RAMA TIRTH Library Education in India and America (Herald of library science, vol. 5, 1966, pp. 152-160; vol. 6, 1967, pp. 138-139).

13. GELFAND, M.A. Survey of University of Delhi, Department of Library Science (ILA bulletin, vol. 3, 1967, pp. 114-134).

14. KAULA, P.N. Survey of Library Science Education in the Banaras Hindu University, 1970.

15. KAULA, P.N. An Evaluation of Education for Librarianship in India. Op cit. pp. 188.

APPENDIX 1

CHRONOLOGICAL TABLE OF THE LIBRARY SCHOOLS IN UNIVERSITIES IN INDIA

Place of School	Name of the University School	Year of Foundation	Name of Course
Madras	Madras	1931	Cert Lib Sc
Waltair	Andhra	1935	Dip Lib Sc
Madras	Madras	1937	B Lib Sc
Varanasi	Banaras	1941	Cert Lib Sc
Varanasi	Banaras	1942	B Lib Sc
Bombay	Bombay	1944	B Lib Sc
Calcutta	Calcutta	1946	Dip Lib Sc
Delhi	Delhi	1947	B Lib Sc M Lib Sc Ph D
Aligarh	Aligarh	1951	Cert Lib Sc
Baroda	Baroda	1956	B Lib Sc
Nagpur	Nagpur	1956	B Lib Sc
Poona	Poona	1956	B Lib Sc
Ujjain	Vikram	1957	B Lib Sc
Aligarh	Muslim	1958	B Lib Sc
Hyderabad	Osmania	1959	B Lib Sc
Jaipur	Rajasthan	1959	B Lib Sc
Chandigarh	Punjab	1960	B Lib Sc
Trivandrum	Kerala	1961	B Lib Sc
Jaipur	Rajasthan	1961	B Lib Sc
Pachmarhi	Saugar	1962	B Lib Sc
Lucknow	Lucknow	1962	B Lib Sc
Dharwar	Karnatak	1963	B Lib Sc
Jadavpur	Jadavpur	1963	B Lib Sc
Bombay	SNDT	1963	B Lib Sc
Ahmedabad	Gujarat	1963	B Lib Sc
Gwalior	Jiwaji	1964	B Lib Sc
Kolhapur	Shivaji	1964	B Lib Sc
Burdwan	Burdwan	1964	Dip Lib Sc
Varanasi	Banaras	1965	M Lib Sc
Mysore	Mysore	1965	B Lib Sc
Rewa	Rewa	1965	B Lib Sc
Gauhati	Gauhati	1966	B Lib Sc
Varanasi	Varanaseya Sanskrit	1967	Dip Lib Sc
Bombay	Bombay	1968	M Lib Sc
Jabalpur	Jabalpur	1968	B Lib Sc
Kurukshetra	Kurukshetra	1969	B Lib Sc
Patiala	Punjabi	1969	B Lib Sc
Bhopal	Bhopal	1970	B Lib Sc
Chandigarh	Punjab	1970	M Lib Sc
Saugar	Saugar	1970	B Lib Sc

APPENDIX 2

STATE-WISE LIBRARY SCHOOLS

S.N.	State	University Schools	Other Schools	Total Number of Schools
		STATES		
1.	Andhra Pradesh	2	1	3
2.	Assam	1	-	1
3.	Bihar	-	1	1
4.	Gujarat	2	1	3
5.	Haryana	1	1	2
6.	Himachal Pradesh	-	1	1
7.	Kerala	1	1	2
8.	Maharashtra	5	2	7
9.	Madhya Pradesh	7	1	8
10.	Mysore	2	2	4
11.	Orissa	-	1	1
12.	Punjab	1	1	2
13.	Rajasthan	1	-	1
14.	Tamil Nadu	1	1	2
15.	Uttar Pradesh	4	4	8
16.	West Bengal	4	1	5
		UNION TERRITORIES		
17.	Chandigarh	1	1	2
18.	Delhi	1	3	4
	Total	34	23	57

Note: Other schools include DRTC, and schools at INSDOC, Polytechnics, associations and colleges.

APPENDIX 3

COURSES FOR SEMI-PROFESSIONALS

Association Schools

1. Andhra Pradesh Library Association
2. Bengal Library Association
3. Bihar Rajya Pustakalaya Sangh
4. Bombay Library Association
5. Delhi Library Association
6. Gujarat State Library Association
7. Indore Divisional Library Association
8. Kerala Librarians Association
9. Maharashtra Library Association
10. UP Library Association (3 Centres)

Universities

11. Rajasthan University
12. Madras University

Government Schools

13. Library Training School, Bangalore
14. Library Training School, Gwalior
15. Library Training School, Simla
16. Library Training School, Madras
17. Library Training School, Patna

Colleges

18. Agra College, Agra
19. BR College, Agra
20. Dharam Samaj College, Aligarh

APPENDIX 4

AN ANALYSIS OF THE FULL-TIME AND PART-TIME TEACHERS IN THE LIBRARY SCHOOLS IN INDIA

School	Status of the Head	Full-time	Part-time	Total
		UNIVERSITY SCHOOLS		
Aligarh	Reader	4	-	4
Andhra	Reader	2	4	6
Banaras	Professor	4	4	8
Baroda	Reader	-	5	5
Bhopal	Lecturer	1	1	2
Bombay	Professor	-	4	4
Burdwan	Lecturer	-	4	4
Calcutta	Lecturer	2	9	11
Delhi	Professor	6	3	9
Gauhati	Reader	2	1	3
Gujarat	Professor	1	2	3
Jabalpur	Lecturer	2	-	2
Jadavpur	Reader	2	2	4
Jiwaji	Lecturer	2	-	2
Karnatak	Reader	3	6	9
Kerala	Professor	3	2	5
Kurukshetra	Reader	2	3	5
Lucknow	Lecturer	2	1	3
Madras	Professor	4	1	5
Mysore	Reader	1	3	4
Nagpur	Professor	-	5	5
Osmanis	Lecturer	2	2	4
Poona	Professor	2	1	3
Punjab	Professor	1	3	4
Punjabi	Reader	2	1	3
Rajasthan	Reader	5	-	5
Rewa	Lecturer	2	1	3
Saugar	Reader	-	3	3
Shivaji	Reader	1	2	3
SNDT	Reader	1	4	5
Varanaseya Sanskrit	Lecturer	-	3	3
Vikram	Reader	5	2	7
		INSTITUTIONAL SCHOOLS		
DRTC	Professor	5	-	5
INSDOC	Reader	1	4	5
		WOMEN'S POLYTECHNIC		
Ambala	Lecturer	2	-	2
Bangalore	Lecturer	3	-	3
Chandigarh	Lecturer	2	-	2
Delhi	Lecturer	2	1	3
Jullunder	Reader	2	1	3
Rourkela	Lecturer	2	-	2

APPENDIX 5

Number of Professionals Required

SN	Type of Library	No. of Units	No. of Professionals of Category A	B	C	D	E	F	G	Total
	PUBLIC LIBRARY SYSTEM									
1.	Librachine	12,067						38,821		38,821
2.	Branch Library	8,089						42,063		42,063
3.	City Central Library									
	A	76					76	866	76	948
	B	23					23	391	23	437
	C	6					6	124	6	136
	D	4				4	4	93	8	109
	E	3				3	3	83	6	95
4.	Rural Central Library	315				315	315	1,890	630	3,150
5.	State Central Library	16			16	32	144	899	320	1,411
	Total				16	354	571	85,230	1,069	87,170
6.	National Central Library	1	1	1	4	147	147	101	120	522
	Total		1	2	20	501	718	85,331	1,189	87,692
	ACADEMIC LIBRARY SYSTEM									
1.	Universities and Similar Bodies	75	-	-	20	55	675	4,215	1,500	6,465
2.	Colleges	2,000	-	-			2,000	46,400	4,000	52,400
3.	Secondary School	70,000							70,000	70,000
	Special Library and Documentation Centre	300	5	5	15	100	200	24,000	3,000	27,625
	Total		6	7	55	656	3,593	159,946	79,689	244,182

PROFESSIONAL DEVELOPMENT IN LIBRARY-INFORMATION SCIENCE AT THE CITY UNIVERSITY OF NEW YORK

Vivian S. Sessions
Associate Professor, University Graduate Division
The City University of New York

SYNOPSIS

CUNY's Professional Development Program is closely linked -- in concept and effectuation -- to the longer-range goal of a Ph.D. in Library-Information Science. The over-all framework is provided by four groupings of courses, of which the most general contains both humanistic (or qualitative) and quantitative concepts The three areas available for more focussed study are information access, its management, and its technology.

Introduction

The interest in Professional Development for Librarians that prompted the establishment of a Center for the Advancement of Library-Information Science at The City University of New York is by no means limited to that institution. The 1970 work by Denton, Between M.L.S. and Ph.D., (American Library Association) is the most important current study in this area. It reveals that several library schools are also concerned about non-doctoral education beyond the Master of Library Science or Master of Science in Library Science. The ALA study also reveals a lack of consistent pattern, if indeed a pattern can be discerned.

One evident fact is the existence of some relationship between post-masters professional development and more formal doctorates. Under these circumstances, it is interesting to note that the program under discussion has been undertaken by the University Graduate Division itself, the administrator of the Ph.D. programs throughout The City University (CUNY). Besides, no other kind of academic unit can have a greater interest in the improvement of those professional information services that support graduate research.

Library-Information Science at CUNY

What Denton did not -- and could not -- include was any reference to the Center for the Advancement of Library-Information Science at CUNY. The reason is quite simple: the Center was being formulated, and its programs subjected to their first live tests, at the very time that the ALA work was being prepared for publication. Furthermore CUNY's Center had, from its first inception as an idea, closer ties to the Special Libraries Association (SLA) and the American Society for Information Science (ASIS). It is therefore interesting that the program now in implementation anticipated -- hopefully successfully -- so many of the problems raised by the ALA study.

It is no coincidence that the program under discussion incorporates a great deal of the thinking that surrounded the Curriculum Committee Workshop of the ASIS Special Interest Group in Education. It is also no coincidence that it draws upon several years of workshop experience by the Documentation Group of the New York Chapter of SLA. The Director of the CUNY Center was very much involved in both sets of activities. The program also draws heavily upon the experience of a predecessor at the University Graduate Division, Project URBANDOC, an R and D effort in information storage and retrieval that was funded by the U.S. Department of Housing and Urban Development.

The background of the Center virtually dictated that its Professional Development represent an integration of librarianship with the computer-inspired developments that have transformed the older term "documentation" into "information science". Thus "Library-Information Science". The earlier project experience of the principal staff (Project Director and Systems Analyst) also indicated that there would be an emphasis on practicality, on the ability to input and output information. However, the sponsorship of the Graduate Division also guaranteed the presence of a strong component of scholarship even though formal academic credits are not presently involved.

Relationship to Doctoral Study

Although the Professional Development Program under discussion was established for the express purpose of helping practicing librarians and other information personnel become better acquainted with newer developments in their field, all the courses relate to another responsibility of the Center for the Advancement of Library-Information Science. That is to develop full-fledged Ph.D. studies in this field. The kinds of work now being given are therefore not necessarily terminal; they may also satisfy some of the requirements for the Ph.D. The only concession to the present somewhat informal status of the programs is that the instructors suggest, but do not require, outside readings, homework, and laboratory experience.

It is anticipated that when the University does authorize the granting of academic credit in Library-Information Science -- hopefully in another year -- it will continue to enable non-candidates to participate appropriately. There are also discussions underway which would widen the non-doctoral students' options to the point where he or she could elect to obtain a second masters degree (or a first, if entering without one). If necessary, the Center is prepared to design a minimum number of dual offerings -- one set for formal credit and the other with lesser demands on the students. That, however, will not be done unless the necessity for duality is proven both pedagogically and financially.

Courses: Design and Implementation

There are presently four areas of instruction in the over-all design. The first is "general" in that it contains those courses, seminars and colloquia that are considered to be more applicable to the entire program than to any one area of concentration. Equally at home in the general category are such humanistically-oriented courses as Information in Society and such quantitatively-oriented ones as Mathematical Concepts. The latter course was the first of this group to be implemented. Successful completion of basic work in set theory, probability and statistics is considered as essential for any serious work in library-information science. It means, for one thing, the ability to read the Journal of the ASIS in its entirety.

The three other areas represent three foci of concern: information access, management, and technology. In a formal degree-granting program, the core curriculum would include at least the introductory work in each; a similar spread is recommended for non-credit students.

1. Information Access

The offerings can be expected to be heaviest in information access, which in this program means chiefly bibliographic access and control. During the initial year of Professional Development, there were three courses in this area: Document Analysis, Document Analysis Workshop, and Introduction to Information Retrieval. The first was strictly a "talk" course, while the other two included interaction with operational information systems. The students in the Document Analysis Workshop prepared a Thesaurus, and then used the Thesaurus to index a selection of NY Times articles. Their input was processed for them, and they received as output two sets of indexes. One used the programs that had been developed by Project URBANDOC and the other was processed by the American Petroleum Institute.

The Information Retrieval class was divided between introductory lectures, the opportunity to address class-formulated queries to four large-scale IR systems (two of them SDI oriented, and two of them offering retrospective searches), and the opportunity to prepare and have processed within the University Graduate Center a small data base of its own. The class also queried its own data base, comparing the effects of searching on descriptors with the effects of searching on author, title, and abstracts.

The Library-Information Science program is particularly fortunate in having acquired a programming package for the IBM 1130 that will permit the students to simulate with small data bases most of the search capabilities of larger computer installations. The resident staff of the Center expects to devote a considerable effort in the coming year to furthering the teaching capabilities of this system. More advanced students will be expected to gain greater interaction with the system itself, whether or not they are candidates for degrees. The capabilities of the system also include substantive data retrieval, one of the many additional units to be added to the information access area.

2. Management

In this area, the Professional Development Program started with Management Science Concepts. It soon developed, however, that this course could advantageously concentrate on the Operations Research alone, which might remove it to the more general group of courses. In any case, the next two courses scheduled for implementation are Financial Administration and the Information Industry. Plans for the still more distant future include a variety of other management-oriented studies in Library-Information Science. The large reservoir of business administration expertise at the Bernard M. Baruch College of the University offers unlimited educational opportunities in this area, with many administrative possibilities for effectuating the opportunities.

3. Technology

The third and last area of focus is the obvious one of technology. The Professional Development Program started with a course in Computer Concepts and Programming. It utilized the PDP-8 computer and terminals, thus starting students early with hands-on experience. Although a small and relatively inexpensive system geared primarily to quantitative analysis, its teaching capabilities for Library-Information Science are by no means that limited.

Both the "hardware" and the "software" aspects of the technology will be developed further, with the IBM 1130 being used for on-site processing as well as remote access to a large IBM 360. However, the prime technological thrust of Library-Information Science is not to develop computer scientists. If specific needs for advanced study in that area should arise, appropriate doctoral programs at the University will be called upon to furnish assistance.

People: Students, Faculty, Advisors

1. Students

The appeal of the Professional Development Program in 1970-71 was entirely to "working stiffs" in libraries and information centers. Their employers, many of whom paid the tuition, range from the army to the university, from drug companies to advertising agencies, from the professional society to Wall Street. A total of 89 individual students (not counting librarians from the City University Graduate Center), participated to the extent of 135 course registrations. Although all classes were scheduled for 14 weekly meetings each, from 6:30 to 8:30 in the evening, a great many of the students travelled from the far reaches of the metropolitan area.

A general characteristic of the student body is that its active involvement in professional activity is by no means limited to this particular program. Practically all the students are members of professional societies, many of them belonging to several. Those societies were, in fact, the chief source of recruits; the University was reluctant in this initial phase of the program to use newspaper or other general advertising which might attract people who did not already have a considerable committment to library-information science.

The program does, however, exclude non-joiners; they report being alerted to the existence of the program by announcements in such places as <u>Library Journal</u>.

2. Faculty

Staffing the program was, of course, crucial. The two resident members of the Center for the Advancement of Library-Information Science were the Director and the Systems Analyst, the first a professional librarian and the second an IBM-trained systems analyst. With over five years of working together on Project URBANDOC, they had integrated library and information science at the working level long before the time to formulate the relationship as a theoretical proposition. There was also some individual and joint experience at the workshop level with the kinds of librarians who later enrolled in the Professional Development Program.

The basic teaching team was supplemented by laboratory assistance from the Graduate Center Computer Facility and by teaching assistance from Baruch College, which houses the City University's doctoral program in business administration.

In addition, the 1970-71 program called upon two leading practitioners in the field: Everett Brenner of the American Petroleum Institute and Melvin Weinstock of the Institute for Scientific Information.

There are, of course, a great many other sources of qualified instructors in the New York area, with several library schools, information systems, and important libraries located in or near the city. The present thinking, however, is to turn more to the various faculties of the City University, and particularly the doctoral faculty which has its headquarters also at the Graduate Center. The present emphasis at the University on interdisciplinary research is especially relevant to library-information science, which can profitably incorporate expertise from both the "hard" and "soft" sciences.

3. Advisors

The potential broad academic base of library-information science at CUNY is already represented by the membership of the Committee for the Advancement of Library-Information Science:

- Professor James M. Beshers, doctoral program in sociology;
- Professor Barry S. Brook, Executive Officer, doctoral program in music;
- Professor Carl Helm, doctoral program in educational psychology and Director, Graduate Center Computer Facility;
- Dr. Richard Logsdon, C.U.N.Y. University Dean of Libraries;
- Professor Margaret K. Rowell, Librarian-in-Charge, Graduate Center Library;
- Professor Vivian S. Sessions, Director, Center for the Advancement of Library-Information Science;
- Mrs. Lynda W. Sloan, lecturer, Center for the Advancement of Library-Information Science;
- Professor Lawrence R. Zeitlin, doctoral programs in business and psychology.

The formal identifications do not, however, properly indicate the extent of individual expertise in library-information science. Professor Brook initiated and directs RILM (Répertoire International de Littérature Musicale). Professor Beshers is an authority on the use of Census tapes, especially for demographic analysis. Professor Helm has extensive experience in computer technology, especially for educational applications. Professor Zeitlin is similarly experienced in management information systems as well as in the relationship between information and decision-making process. Dean Logsdon and Professor Rowell are particularly experienced in the areas of academic and research information needs and resources.

The Future: Prospects and Caveats

This is one situation in which it is completely accurate to say that a program enjoys the whole-hearted support of the administration. Dr. Mina Rees, President of the University Graduate Division, and Dr. Harold Proshansky, the Dean, not only support professional development in library-information science, they made it possible. Having made available the necessary Graduate Division personnel lines to establish the Center for the Advancement of Library-Information Science, they asked only that the incremental cost of its external teaching responsibilities be borne by student fees. As indeed it was during the first year, and is anticipated for the next. This degree of committment contrasts sharply with the impression of Danton in the ALA study that most post-M.L.S. programs are closely linked to the availability of federal funds.

The immediate prospects for continued and expanded professional development at CUNY are therefore excellent. As for the longer range one of the Ph.D., it is being pursued with all the speed that is possible under the administrative constraints of a large public institution.

In the meantime, some words of caution are necessary to temper the predominately optimistic tone of this report. They have to do with the "science" part of library-information science. The thrust of the entire program appears more in line with the methods of the "hard" sciences than of the humanities that have hitherto dominated the library part of library-information science. Whole semesters of rigorous courses, seminars and laboratories are undoubtedly appropriate to the doctorate, and even to the masters' degree. However, it is possible that non-credit students will be less inclined to push as consistently for proper mastery of their materials. The question then arises as to how the University should treat their requests for some formal recognition of attendance.

Another set of caveats concerns the proper teaching of the courses. Although those that received specific mention above are reasonably straight forward in their staffing requirements, others will require more than one instructor, and/or more than the standard timing. The full implementation of Professional Development in Library-Information Science at The City University may, therefore, require some decisions that other institutions are not prepared to take.

The Director of the CUNY program is very grateful for the opportunity to explore both the prospects and the caveats. Hopefully, the results will make contributions beyond the institution which is supporting the effort. There have already been many requests for copies of teaching materials, and these will be honored as soon as possible.

Reference

Danton, J.P. Between M.L.S. & Ph.D. A Study of Sixth-Year Specialist Programs in Accredited Library Schools. Chicago, American Library Association, 1970.103 p.

A SYSTEMS APPROACH TO EDUCATION AND TRAINING
IN INFORMATION SCIENCE AND TECHNOLOGY

Herbert Schur
Senior Lecturer,
Postgraduate School of Librarianship and Information Science,
University of Sheffield

SYNOPSIS

In the systems approach the detailed design and implementation of a programme follows consideration and discussion in depth of utilitarian and educational objectives and the assignment of priorities, of environmental and other constraints, and of qualitative and quantitative criteria for evaluating the efficiency and effectiveness of the education and training programme. These and other factors are discussed.

Introduction

The systems approach[1] is a method of problem solving which, viewing a system as comprising interacting and interdependent sub-systems of men, machines, materials, and money and itself a sub-system of a wider system, aims at achieving the overall objectives of the system in the most effective and efficient way. The method includes first and foremost the consideration of the objectives of the system and their re-examination in the wider context, the consideration of constraints, and of criteria. It is the imposing of the requirement for thinking in terms of the wider context that distinguishes this approach from the all too common narrow and parochial approach the consequences of which may be very serious. To find a few examples of a failure to apply a systems approach in much of industrial planning we need to go no further that the polluted beaches, to breathe the polluted air, and to see the derelict land in almost any of the advanced countries. A failure to apply the systems approach in programme design may result in equally serious consequences, in terms of mis-employment, under-employment and unemployment.

Assuming, then, that the systems approach may be of value in educational planning, we still have to ask whether education and training is required in information science and technology. The fact that most of the information workers and researchers do not possess formal educational qualifications in information science or technology and that many of them are successful may suggest that formal education and training in this field is not necessary and that most, or even all, that the information worker or researcher needs to know can be learnt 'on the job'. An analogous situation arises in management education and the case for an academic educational programme in that field, made by Professor Eilon[2], applies with equal force here. He stated: "Let us face it:

some people go through an engineering college and never become good engineers; others become first class engineers without any formal education. Similarly, I do not think that the question whether management education can produce good managers or not is a relevant one. The question is whether it can produce <u>better</u> managers. Naturally you must have a flair for management and certain personal attributes, as indeed for any subject or profession you choose as a career, and without appropriate pre-requisites it is doubtful whether any educational programme can guarantee to produce good managers, good engineers, good doctors or indeed good practitioners in any field of human endeavour. But given the aptitude and inclination for a certain subject, surely the individual stands to benefit from a course which provides a broad background, which attempts to knit together relevant problems and issues and which relates past and present human experience in the field, rather than to struggle on his own and learn from his slowly accumulating narrow experience ".

<u>Objectives</u>

The primary objectives in education and training programmes are educational. It is only when we consider the programmes in the wider context that we have to consider the problem of education and training for what and the implications of these needs for the programmes.

Following Bloom[3], <u>educational objectives</u> may be grouped into three domains:

(1) cognitive,
(2) affective, and
(3) psychomotor.

The cognitive domain, which will be considered in more detail below, includes those objectives which are concerned with the recall or recognition of knowledge and with the development of intellectual abilities and skills.

The affective domain is concerned with attitudes and values.

The psychomotor domain relates to the manipulative and motor-skill area and is therefore largely of interest in the design of training programmes at the technician level.

Returning now to considering the cognitive domain, we have, in order of increasing intellectual difficulty, six categories:

(1) knowledge (which includes the recall of facts and generalizations, methods and processes, patterns and structures),

(2) comprehension (which includes the ability to make use of the material or idea being communicated without, necessarily, relating it to other material or seeing its fullest implications; it includes the ability to translate, interpret and extrapolate from facts or ideas),

(3) application (which includes the use of abstractions in particular or concrete situations),

(4) analysis,

(5) synthesis, and

(6) evaluation (including judgements in terms of internal or external evidence).

In analysing existing programmes and courses within these, it is difficult to detect just what stress is laid on the various categories (for instance by studying the examination papers) since solving a problem encountered for the first time may require the highest intellectual ability, yet if the problem and its solution had been encountered earlier, only simple recall may be involved. The need for educating and training students for work in a dynamic situation, rather than a static one, means that a right balance must be found between what Henderson[4] called "knowledge-stuffed memorizing-based type of teaching" and the opposite extreme where a lack of factual knowledge may lead to the "re-invention of the wheel". Whilst much research is concerned with the re-examination of "facts", progress would not be possible if every problem had to be considered *ab initio* and if all, or even a large part, of accumulated knowledge were to be ignored in education and training. There is thus a need for a student to acquire a knowledge of some basic facts, generalizations, methods, processes, patterns and structures regardless of whether he is to proceed to research or to operational work in an information system. The problem is to select the bits of knowledge required to be taught (and therefore identifying the large areas of knowledge in which the bits required can be acquired by the student at a later stage, for instance in the course of his research or operational work). Since the amount of knowledge which can be acquired by a student during the period of formal education and training is very small indeed in relation to the total knowledge of a field, to attempt to cram more and more facts and techniques into a programme will only lead to a massive indigestion on the part of the student and yet make barely an impact on the amount the student is ignorant of; the aim should therefore be to ensure that the student learns to recognize the situations when he does not know and to ensure that he does know how to go about acquiring the required knowledge.

The second of Bloom's categories is comprehension. This, as stated above, includes the ability to translate, interpret and extrapolate from facts or ideas. Translating includes the expressing of facts or ideas stated in one set of words in other words (both between different languages and in the same language, as when attempting to express ideas in non-technical terms), it includes the translating of mathematically expressed statements into verbal ones, and *vice versa*, and of graphs into mathematical or verbal statements, and *vice versa*. It includes interpreting, for instance of instructions for the use of classification or indexing schedules or of data presented in numerical table form, and extrapolating from data presented in various forms. Do we really make sure that students are capable of doing this? My impression is that students who specialized in the physical sciences or mathematics all too often assume that once, say, experimental results are presented in a mathematical or graphical form anyone will be capable of comprehending these results without needing to be told in words; on the other hand humanities background students will write at great length even where a graph or a mathematical formula would express all they have to say in a concise form. It is not only a matter of comprehending facts and ideas expressed in an unfamiliar form but also of an ability for extracting the full content of the statement.

Closely related to comprehension is application. Here we expect the student to search for familiar elements in a problem, to use these to restructure the problem, to select the theory, principle, idea or method suitable in solving problems of this type, and finally to use these in producing the solution. To take a simple case: we wish to know how to satisfy a national loan demand for scientific periodicals with the minimum number of multiple sets of these periodicals; that is, would it be better to have decentralised regional collections or a centralised national one? The student, by making assumptions such as that the loan demand for any given volume of a given periodical will follow a Poisson-type distribution, and that the cost of postage and packing is the same for the regional or national service, can then produce figures showing the number of sets of the periodical required for various loan demands and from these argue the case for, or against, a centralized or a de-centralized service. The point here is that more than just comprehension (e.g. of terms such as "Poisson distribution") is required. Programmes exist which include a wide range of advanced topics, one 20-hour (in total, not per week) course, with 40 hours of 'practicals' including, *inter alia*, methodology of research in information science and technology, graph theory, Boolean algebra, information theory, operational research, psychology of classification, semantic theory, linguistic models, transformational grammars, and a lot more besides. One may be a bit sceptical of the ability of students to comprehend all this in the time devoted to it, let alone in applying it all. This is not to say that such topics should not be included in a programme, but only to point out that mentioning a wide range of topics is not sufficient. Time has to allowed for carefully selected topics to be studied in depth, so that the knowledge gained can be applied.

An educational programme, particularly for higher-level students, should ensure that analysis is given sufficient weight. Before solutions can be proposed, problems need to be analysed, and this applies equally to operational work and to research. But analysis, and by this I mean the breakdown of structures (e.g. systems), processes or statements into their constituent parts and the critical and logical study of the parts and the whole, should not be considered as an end in itself, but should be viewed as the prelude to further work, not necessarily of synthesis. Statements like "information is a raw material" or "information is a commodity" should be analysed, as should such very difficult concepts as "relevance", or "information needs" of users. There is a large number of topics on which useful analysis can be carried out and even though no definite conclusions may be reached (e.g. on the meaning of "information", on appropriate measures of "benefit" of an information system), the analysis of the problems will in itself help to clarify at least the areas of our ignorance.

Whilst analysis enables the students to dissect, constructive work requires synthesis. It is synthesis which appeals to the imaginative mind and it is this imagination which should be fostered in an educational programme. Developing a capability in synthesis may be achieved by means of system design studies, by report writing, by planning exercises, and other means. It is the development of this capability which appears to be most neglected in schools and even in the undergraduate courses in universities and it is this which was stressed in various major reports on engineering education[5] and in my recent discussions with deans and professors in some of the leading schools of information science in the USA and Europe.

Finally, evaluation. This includes judgement in terms of internal or external evidence involving the use of logic and estimation of probabilities (among others). In general, scientists, and particularly physical scientists, are used to judging on the basis of data which are on at least the ordinal scale and, perhaps because of this, appear to be unhappy when asked to make value judgements as in some major management decisions. It is possible to teach how to evaluate and judge on the basis of objective (and quantitative) data, but such exercises remain 'theoretical' if the student bears no responsibility for the consequences of his judgements. The student who will cheerfully decide that an amount one hundred times his annual salary be spent on a machine for the hypothetical information service, will, when later required personally to be responsible for sanctioning an expenditure one tenth of his salary feel weighed down by responsibility. Can 'games' really simulate real-life situations? Nevertheless, 'games' may be of value in bringing out some of the problems of making judgements and thus prepare the student to face real-life situations forewarned and forearmed.

Professional objectives

Since education and training does not take place in a vacuum, objectives of the programmes have to be considered in the wider context. In systems language, the programme becomes a sub-system of the larger system of education and training in the country, which itself is a sub-system of national policy, and so on.

In determining the professional objectives of a programme, usually at least four interests need to be taken into account, namely those of:

(1) the individual,
(2) academic,
(3) professional, and
(4) national.

The individual (i.e. the student) wants the best possible education and training programme for his needs. But what is 'best' and what are his 'needs'? The individual is usually singularly ill-prepared for answering these questions and therefore decisions may need to be taken on his behalf by those with experience in the field. I would think that in designing a programme consideration must be given to the career expectations of the graduates, which must not be worse than the expectations of graduates in related fields (e.g. science, engineering), in terms of financial reward, prestige, and interest in the subject. We must avoid developing programmes leading to dead-end jobs, however great the demand from 'the field' for such personnel, and, on the contrary, we should aim at providing an education for flexibility. Education for flexibility will not only increase the employment opportunities for the individual, but also by exposing him in the course to completely new topics will broaden his outlook and, I would hope, increase his interest in the subject. There is conflict here between the short-term and long-term benefits of a programme for the individual, between the benefits to the individual and to the employer, and between those of the employer and the state. So long as the conflict cannot be resolved by an appeal to objective evidence, an essentially subjective decision has to be taken, and, in the circumstances prevailing in the U.K., I give considerable weight to what I believe are the needs of the individual. However I am not so sure that an equal weight would necessarily be appropriate in, say, a developing country or Israel.

The academic objectives are concerned with promoting fundamental theoretical and experimental studies. From the point of view of academics, advanced level programmes need to provide adequately educated and trained candidates for research work, and therefore programmes must include the necessary foundations.

Professional institutions are concerned with maintaining a high level of quality in those entering the profession, the level being similar to that demanded by professional institutions in related fields. The usual set of requirements includes a prescribed level of general education, of the specialized education, and of practical experience. It is usual for the major professional institutions to recognize certain academic institutions, or particular programmes, for the purposes of giving full or partial exemption from the professional qualifying requirements of the institution and it is obvious that a programme which is recognized by a major institution will attract more good candidates than one that is not. In planning programmes it is therefore useful to note (but not necessarily follow) the requirements of the relevant professional bodies.

Finally, national objectives. The primary objective is, of course, survival and linked with this is the economic, social and cultural development of the country and nation. As already mentioned, there is obviously a source of conflict between the objectives of the individual and those of the state, but in an honest (i.e. not corrupt) democracy the gap separating the objectives is not likely to be unbridgeable. To illustrate: sources of conflict may be created by the broad versus specialized programme requirements, education versus training, requirement for operators versus research workers, short-term versus long-term needs, and so on. In a democracy the state can only exert positive pressure, for instance by providing funds for the support of programmes which are considered to be in 'the national interest', it cannot prevent a programme from being implemented (although lack of financial and other support may make it difficult to maintain such a programme). Again, the implication for the programme designer is that national objectives should be noted but that their 'official' interpretation should not, necessarily, be accepted.

Categories of work

So far I have been discussing education and training programme objectives in general terms. I now wish to turn to a discussion of the various categories of work.

As we had already stated earlier[6], a useful division can be made into four categories, namely:

(1) production and operation,
(2) development and design,
(3) applied research, and
(4) background research.

In considering the design of programmes in relation to the requirements of the production and operation category of work, as well as of the other categories, the first question we have to ask is "what is the business we are in?". Does information work include, for instance, all aspects of data processing including automatic equipment? It is concerned with "all the phenomena and processes mediated by signs, the carriers of information", or is it concerned primarily with the "collection, collation, evaluation and organized dissemination of scientific and technical information" only? The definitions of information science and technology differ widely, but perhaps a description of the field is emerging in publications such as the Annual review of information science and technology, periodicals such as Information storage and retrieval, Nauchno-tekhnicheskaya informatsiya, Journal of documentation, Journal of ASIS, and abstracts such as Library and information science abstracts, Referativnyi zhurnal: n.t.i. and its English version. The production and operation category will require the largest share of the output of education and training programmes in information science and technology, and that at a range of levels.

I believe that development and design is a key activity in all but the smallest operational units and therefore the requirements of this activity must be taken into account in the design of programmes.

Applied research is required both to back-up the development and design effort and to provide the means of bringing the findings of background research to the stage when these can be used in design. The applied research worker will need to comprehend the background research findings and analyse and synthesise these for the benefit of the design workers.

Finally, the background research, whilst fairly remote from the problems of day to day operation of information units, is required to provide the scientific basis for the development of new methods and means in information science and technology. It will necessarily be closely linked with work in related fields, such as psychology, linguistics, communication studies, and engineering science.

Priorities in categories of work

Looked at from the point of view of national needs it is obvious that the first priority will be for operational staff. What is not so obvious is that even in a developing country some staff will be required for development work since information systems can remain unchanged only at the risk of becoming increasingly ineffective, inefficient, and obsolescent.

I would rate design capabilities as a close second priority. Adapting systems imported from abroad may be useful in a high proportion of cases, but there will always be particular local needs for special purpose systems which are best designed by one with an intimate knowledge of the local conditions and requirements. I am referring here to the design of systems not only in cases where security considerations prevail, but also of systems such as district or national library and information systems.

Applied research may be conducted on an ad-hoc basis, that is whenever the need arises for providing background information for designers, or on a more systematic basis in accordance with some policy. The piecemeal approach, which is essentially based on trouble-shooting, cannot provide a basis on which education and training programmes can be planned, and therefore a national policy should be evolved designating certain areas as priority areas. These priority areas would include first those areas of interest to the country on which little or no research is being carried out elsewhere (for instance problems connected with the language barrier[7] and language processing), secondly problems specific, or of primary interest to the country (problems connected with specialized information centres), and only thirdly problems of general interest in information science and technology. This may, of course, not be a popular view since the individual research worker may gain more prestige internationally by working in areas of more general interest, but the consequence of this attitude will be that the local needs will be neglected and design and development will suffer. If priority areas are known then centres concerned with their study can be developed and consequently education and training programmes based on the expertise in these centres can be planned.

Background research, on the other hand, appears to me to be essentially 'academic', and although possible ultimate applications may be imagined these applications are very remote and the research is primarily oriented towards an increase of knowledge of the field rather than the potential application of this knowledge. It is not really feasible to plan background research, both because the number of powerful original thinkers who are able and willing to devote themselves to background research is very limited, and because original thought cannot be planned. The best that can be done is to attempt to promote interest in information science and its problems and to encourage those willing and able to work on these.

Even if it is decided to support education and training for categories other than operation, there is no simple way of determining what proportion of the total effort will be required for this support. It may therefore by useful to note that in engineering, much of which originated in craft-based industries and not in research laboratories, and much of information work practice is still in the craft stage, the proportion of the graduate manpower engaged in research and development has reached about 35-40% (in countries like USA[8] or the UK[9]); it is therefore not unreasonable to expect that a proportion of a similar order of magnitude will be required in information science and technology by, say, 1990.

<u>Summary</u>

We have, so far, asked and outlined answers to questions such as:

(1) why do we need education and training programmes?
(2) if we do need such programmes, what are the educational objectives requiring to considered in the design?
(3) and what are the objectives the individual, the academic community, the professional community and the state wish the programme to achieve?
(4) there exist various categories of work: what are the principal ones requiring to be considered in the design?
(5) and what are the priorities in providing for education and training for these categories?

<u>Constraints</u>

Programmes are implemented not in some ideal environment, but in a real one. We therefore need to take into account the constraints imposed on the programme design by various factors. In parenthesis, it should be stated that, as far as I am aware, no useful studies had been made of the sensitivity of the effectiveness and efficiency of programmes to these constraints (e.g. what is the effect of a lack of material facilities, of good teachers, of different educational systems, etc.).

Let us consider some of these constraints:

(1) the HIP barrier (i.e. history, incentives, psychology)
(2) educational system
(3) students
(4) teachers
(5) material facilities, and
(6) finance.

The HIP barrier is by far the most important one for unless there is the will and determination for change no action will result. Following Galloway[10], the '<u>history</u>' element of the barrier includes not only the history of the individual person concerned with the change, but also the history of the institution, of the profession, and of the country. All are greatly affected by their past and by the effect of this on their attitudes to change, and particularly rates of change. In practical terms one consequence might well be the need for developing a completely new programme rather than attempting to modify an existing one.

Change can be brought about, or speeded up, by general exhortations, by the 'carrot' and by the 'stick' method. General exhortations and resolutions, so popular with some governments and international bodies, are of little value[11]. The 'stick', that is penalties and sanctions, may produce the desired results but at a considerable loss of good will and enthusiastic cooperation. We are therefore left with what, in my view, is the most desirable method, that of the 'carrot'. <u>Incentives</u> are: survival (both national and individual), economic well-being, prestige, and interest in work. Some of these incentives have already been discussed in relation to objectives, and it only needs to be added that a well designed national programme of education and training in information science and technology will make a positive contribution in all these areas.

Psychology is the third element of the HIP barrier and is the one least written about in accounts of programmes. Nevertheless it is a very real factor and cases abound where progress had been held up because individuals would not co-operate for reasons such as imagined (or even real) 'loss of face', or loss of prestige. Obviously inside knowledge of the likely reaction of personalities to be involved in and affected by any proposals considered for implementation in the programme is of great importance since this can be taken into account in the design. It is not only what is proposed but also how it is proposed which matters.

The second major constraint is the educational system which determines what is acceptable where, for whom and when, and how it is to be implemented. If a programme is to be implemented in a university, for instance, then the design must be such that the programme is acceptable in its level, form, and other parameters to the institution concerned. To the extent that educational systems differ between countries (consider just the place of engineering education in some of these), to that extent will vary the assumptions underlying similar level programmes in these countries.

The principal constraints relating to students are their intelligence, educational background and practical experience, motivation, age, and numbers. Admission requirements to most existing programmes specify some, or all, of these although, so far, no figures appear to have been published to show the sensitivity of programme effectiveness and efficiency to variations in these requirements. There exist, however, some figures in relation to undergraduate programmes in a university12, which indicate that the relations between academic progress and factors like intelligence or educational background are not quite as simple as might be expected from the rather rigid entrance requirements specified for some programmes. On the other hand students with a relatively low I.Q. and a poor record of educational progress are much more likely to fail than their fellow students. Similarly, studies have been carried out (although not in the field of education and training in information science and technology) of effects of motivation, age and numbers (in seminars, in a university as a whole), on success in programmes. The results of these studies do, of course, indicate the importance, or otherwise, of these factors, but for programme design and implementation more research into these factors would be desirable.

The shortage of suitable persons who are able and willing to teach is, in many cases, a serious obstacle to the development of programmes, particularly those for higher levels. The design of programmes must therefore take this factor into account and build from the available to the desirable on the basis of the attainable at the time the programme is being planned for. To wait until the whole programme can be implemented by first class teachers means to postpone the implementation indefinitely and this, in turn, means to destroy the programme, since objectives, constraints, and criteria are bound to change in a dynamic environment.

Similar considerations apply to the availability of material facilities. These facilities often exist, or can be built up during the planning stage provided that the requirements are realistic. In most countries fairly large computers are available and collections of the literature of information science and technology can be built up to complement existing collections of academic and technological literature. Here I would like to stress, again, the importance of periodicals relative to textbooks, particularly for programmes for the higher levels. Here, too, there is considerable advantage in implementing programmes in existing educational institutions (or some large research organisations with educational facilities) with their large libraries and other facilities.

Finally, finance. The provision of financial support is crucial to the planning and implementation of programmes. Both the total amount and any restrictions on the expenditure need to be taken into account in programme design. This applies not only to the provision of funds for the build-up of material facilities and for remuneration for the teaching staff, but also the provision of finance for the support of students. From the national point of view, if research and development are considered desirable then the provision of information services is desirable, and if this is so qualified staff must be desirable. This implies that their education and training should be considered as of equal importance to the education and training of research and development workers (scientists, engineers, economists, etc.).

This brief sketch of some of the constraints shows that even for identical programme objectives different solutions will require to be found in different countries, or even in different regions of the same country. This does not of course mean to imply that every component course of a programme will need to be designed *ab initio*, but if it is to be taken over ready made a suitable one will need to be selected.

Criteria

There appear to be potentially three useful indexes for appraising education and training programmes:

(1) effectiveness
(2) efficiency, and
(3) benefit/cost.

When asking whether or not a programme is effective, we are asking whether or not the programme is achieving what it was designed to achieve. If we ask, "how effective?", we are attempting to determine the ratio of actual achievement to planned achievement.

We have earlier distinguished between (i) educational objectives, and (ii) professional objectives. Effectiveness indexes can be determined provided that we can quantify the objectives. Considering first the educational objectives it appears likely that a series of assessments of the students, by means of examinations, seminar performance, course work, laboratory work, and minor research studies, might provide a reasonable basis for an educational effectiveness index. Whilst it is true that all these assessments are subject to error, as studies of marking of examination papers have shown, it is still the rule that students' performance is assessed and (at least in British universities and professional examinations) the use of external examiners does tend to ensure a reasonable uniformity of standards between universities and between professional institutions.

To quantify professional objectives is considerably more difficult. Is the programme to be considered 100% effective if all its graduates find posts in the area of information science and technology for which the programme was to prepare them, or are only those graduates to be considered who 'survived' a given 'probationary' period in employment (excluding those leaving for reasons of health or for family reasons)? Are only those who achieved a predetermined level of seniority to be counted? Is salary relevant? And publications?

When asking, "how efficient is the programme?", we are attempting to find an output to input ratio. We may, for instance, use the ratio of graduates to students at the start of the programme, or the pass rate, or some similar ratio relating to students. We may, on the other hand, have a mixed ratio like graduates to cost of programme, where the cost needs to be defined closely. To amplify: cost may include all the inputs, such as manpower (teachers, administrative staff, and the marginal cost of other persons, plus the cost of students, including their loss of earnings), materials and machines, and overheads including depreciation (an item which may be particularly heavy if one includes depreciation of software and investment in programme preparation). The graduate to cost of programme ratio has also been used in cost-effectiveness analyses[13] either to find the cheapest means of accomplishing a defined objective or, finding the maximum output for a given expenditure.

The benefit to cost ratio requires, again, a definition of the terms used: 'benefit' and 'cost'. Whilst 'cost' is often defined in the sense of input costs, as indicated in the preceding paragraph on efficiency, 'benefit' is still subject to much discussion. The importance of this ratio is that given a limited amount of money available for expenditure on all kinds of necessities (or even luxuries), a decision has to be made on how to spend this money to best effect. Would the same investment in promoting (say) other programmes bring greater benefits? And what would be the losses if the programme were not implemented? Benefits and costs, to whom?

A common approach is to use a 'rate of return' on investment, that is the return to programme graduates in terms of the total lifetime increase in earning resulting from the programme as compared with the cost of the programme. The financial benefits are not the only ones, but other than financial benefits are difficult, if not impossible, to quantify. There are, of course, considerable difficulties in calculating the lifetime increase in earnings, firstly because programmes in information science and technology are of too recent an origin and therefore none of the programme graduates has worked for 30 or 40 years following graduation from the programme, and secondly it is difficult to estimate what the individual might have earned had he not participated in a programme, that is, estimating the degree to which students opting for programmes in information science and technology are representative of a larger body of students attending programmes, in related fields (e.g. science, technology, or economics).

The benefit to cost ratio is probably best used not in comparisons of investments for different purposes, but in comparing programmes aiming to achieve the same, or similar, objectives.

Conclusion

In this paper we considered three primary components in system design of programmes of education and training in information science and technology. We have, intentionally, left out consideration of designing particular programmes and therefore we have not discussed methods of forecasting needs and levels of work (which are likely to be determined by analogy with related fields).

Following a brief note on the need for education and training programmes, we considered both educational and professional objectives (and briefly discussed categories of work and the priorities), a variety of major constraints, and, finally, some of the criteria for assessing education and training programmes.

We hope also to have indicated some of the areas in which more research is required if, in future, programmes are to be designed on a more scientific basis than in the past.

VI Selection, Education and Training of Personnel

References

1.	JENKINS, G.W.	The systems approach, Journal of systems engineering (University of Lancaster, England), Autumn 1969 1 3-49
	OPTNER, S.L.	Systems analysis for business and industrial problems solving. Englewood Cliffs (N.J.), Prentice-Hall, 1965
2.	EILON, S.	Reflections on management science. (Inaugural lecture). London, Imperial College of Science and Technology, 1964
3.	BLOOM, B.S., et al.	Taxonomy of educational objectives: the classification of educational goals. Handbook 1: Cognitive domain. New York, David McKay, 1956
4.	HENDERSON, A.D.	An educator's view of the Goals (of engineering education) Report, Journal of engineering education, 1968 59 (1) 27
5.	WALKER, E.A., et al.	Goals of engineering education: Final report of the goals committee, Journal of engineering education, 1968 58 (5) 376
6.	SCHUR, H. and W.L. SAUNDERS	Education and training for scientific and technological library and information work. London, Her Majesty's Stationery Office, 1968. p. 42-43
7.	HUTCHINS, W.J., et al.	University research and the language barrier, Journal of librarianship, 1971 3 (1) 1-25
8.	---	ref. 5, p. 404-405
9.	---	Statistics of science and technology, 1970 p. 116
10.	GALLOWAY, D.F.	The technology, economy and philosophy of improving productivity in the manufacturing industries, Production engineer, 1968 47 324-329
11.	GALLOWAY, D.F.	ibid., p. 310
	KAISER, H.	in Information und Dokumentation in Wissenschaft un Technik, GdT (Gemeinschaftsausschuss der Technik) Schriften, 1969 (1) 32
12.	SCHONELL, F.J. et al.	Promise and performance: a study of student progress at university level. London, University of London press, 1962
13.	BLAUG, M.	Cost-benefit and cost-effectiveness analyses of education in Budgeting, programme analysis and cost-effectiveness in educational planning. Paris, Organization for Economic Cooperation and Development, Directorate for Scientifi Affairs, 1969. p. 181-184

THE NEED OF HIGHLY QUALIFIED LECTURERS FOR THE TRAINING OF INFORMATION OFFICERS

G. Reichardt
Head of the Central Library of the Nuclear Research Center Jülich

SYNOPSIS

Besides all new programmes for the training of information officers, well experienced lecturers are necessary for each training. The lecturer must be able to inspire his hearers so far and so much that they take enough with them, ready to attack new and better ideas and to become active during the whole of their professional life. Consequently post-graduate seminars for lecturers should be regularly given in all countries of the world. An educational system for lecturers in library and documentation services should be organized at once.

When you invited me to lecture I was quite sure that I wouldn't prefer a common-place subject such as education in Germany or in the town of Jülich or in our Nuclear Research Center. I didn't wish to pay homage to local patriotism that didn't seem to be reasonable. I have had experience in the training of documentalists at Frankfurt and in the training of scientific librarians at Cologne. I believe that many an education programme of today is often non-effective, obsolete, filled with ballast. The basic concept that documentation and information sciences are applied sciences and social science, is missing. Therefore, there is no education for the performance of activities beyond the reach of libraries as such.

That wider reach means: teaching methods, how to advise, to provide, and to satisfy the individual user? That wider reach means: lessons about the various means of discovering unknown literature, often representing secret literary paths for the astonished user; lessons about methods, how to encourage the reader to read the same book as oneself; lessons about the assistance a documentalist should and could give to each visitor wanting to learn and to all those entering a library - the documentalist has to be trained in interhuman relations.

Therefore, we do not often ask: where do lecturers come from? Where are well experienced teachers?[11]. The problem, how to manage information is rarely recognised[9] but education methods are not discussed at all.

Today's lessons in librarianship and documentation are characterised by the features of a sort of pre-school training: educational literature about training in librarianship or in documentation activities presented in form of books, lectures, programmes, conference papers, or in other endless masses of papers, and that hardly any more has meaning than business forms or the bureaucracy of documentation work: library and documentation work are represented in the form of a one-sided conversation.[5] The lecturers seem to look at the students like married spectators watching a wedding and saying: "Let them too see how life is!"

Should education be regarded as successful when the level of knowledge has reached a certain mark of success in the same way as the calculable successes in the production of refrigerators, motor-cars, and other industrial products? By no means!

For a good lecturer must not only take into consideration the human and social relations of libraries and documentation centers during his lectures, but also must point them out. Without establishing these relations for the audience, for students, for users, we could abandon all our documentation service, catalogues, book collections, and bibliographies - and last but not least - ourselves too. I found this fundamental attitude best and most clearly applied by the University of Pittsburgh, Pennsylvania, in its educational programme for Information Science, that attributes great importance to the "Foundations of behavioral theory."[3]

Where are the difficulties coming from which our social structure has put up against a real communication of information? Actually, these are difficulties which the Enlightment has had to overcome in all countries and in various manifestations since 200 years. You find the idea expressed in a popular German proverb: Better the devil you know, than the devil you don't know! This proverb as applied to communication of information means: It is not absurd: The person truly responsible for the efficiency of information is not the one who informs but that one who is informed.[6] According to that discernment the information process is determined as an event, where the receiver, the student, the auditor is the important partner. To request an improvement of educational methods means understanding that documentation work and librarians' knowledge do not unfold themselves like a magic flower, but it means, that in the long-known educational and didactic qualities, requested by Pestalozzi, Comenius, Kerschensteiner or Ellen Key, Eduard Spranger and Anton Semjonowich Makarenko, should also be required for studies of librarianship and documentation: viz. enthusiasm for the subject, patience for the audience, continuity, prosecution of the fixed aim, dynamic formation of the lessons by adding intellectual stimulus without which students respond poorly to lecturers. "Information obtains much, but encouragement obtains all."[1] The teaching of the subject only represents a cliche without the students' own participation. In that way we train only engineers of catchwords. Soon after the examination, all subjects taught fall off the walls like inferior house-plaster.

In general: Documentalists are trained with an orientation towards forthcoming examinations, but actually documentalists oriented towards practice ought to be trained.

For a good lecturer must be able to diagnose, to prognose, to judge the situation in order to lecture and to recommend the right therapy, the right decisions, the right means. The only real method of modern education is exclusively the right analysis of aim and means; that means, an analysis of the professional situation at libraries and documentation centers and an analysis of literature sources. This education, called "training on the job" shall prepare the students to take over a special position and to fill a special function. This analysis when carried out as deliberation and decision should result in a system of relation between teacher and student, the method of which can be shown at a pragmatic model implying good and critical relations between teacher and student and including a good feedback to the teacher, so that the ideas of library and documentation considered to be fundamental and obligatory may be submitted to modification and correction.

During library and documentation lessons we are learning to answer the questions: How? How to proceed? How to analyze and how to catalogue? How to find literature for a special problem? But where are we taught to ask: Why shall we act in a certain way? What does this card index actually accomplish? Which rules and regulations are applied to the administration of the branch library, i.e., of the branches of institutes and laboratories? What is the matter with the basic strategies of cataloguing?[13] Why is that place of employment saved? Why is this colleague preferred and that one dismissed? Where can we learn to be more than a well turning small wheel within the machinery of an institution of industry or research? Where can we acquire the knowledge required to ask, to think, to judge in a self-reliant way? Where are we taught to recognize the margin of action of a documentalist and to assume responsibility for ourselves and others? Where can we learn that it is indispensable to know for each information the relation between the various branches of knowledge involved, whose theories and methods have become extremely complicated? Where can we learn that language of science has become an artificially constructed language based on formalism and symbolism, where a development of pragmatic-semantic aspects results in purely syntactical questioning? We must learn to reach decisions, according to a well prepared planning, to prognosticate, to prepare decisions according to need.[12]

We should encourage the inclusion of studies of the operation, economics and use of existing services and of the need for further development. The training programs should consider the management problems of libraries and information systems and should identify the objectives of future planning and operation and the types of information that are required. The students should learn something about technological forecasting, about PPBS, about analysis of needs, about the mechanics of decision making, about R&D programming and goal determination. The educational reformation must start at this point. The practice of a lecturer will have to assume information about the students' interest in life and about their expectations concerning education as well as a critical opinion and a clear understanding of the present problems. How to educate such documentalists? Where are appropriate lecturers to recognize and to solve those problems and to assume conscious responsibility? Consequently there is a correlation between the request for experienced teachers and the efficiency of education and there exists the same difficult question of the efficiency of lectures. For the lecturer must be able to inspire his students[7] so far and so much that they will become good information officers[9]: that he will be able to establish communication from brain to brain.

The subject of the lectures should be based on the need for things which should be done in library and documentation world, and accepted as their mission by the students themselves. This knowledge may consequently be derived from two main sources:

Firstly: More attention should be paid to the audience, to the students and to the users of information! That means more studies, more research and more lecturing about the environment of library and documentation center.[4]

Just as industrial production can't be stimulated without marketing, library and documentation center, too; and education in librarianship and documentation should take into consideration, more than until now, the environment, the audience and the user that means i.e. = students and their future duties.

Secondly: I don't know whether our time is better than the past or the future, but in spite of all controversies in this regard it is marked by one evident advantage: it admits criticism.

And concerning the methods - we should seek increasingly to identify those methods of study that can best provide the highest type of training and that can stimulate the application of proven techniques, data, and knowledge. Knowledge itself occupies the second position. At this moment, when for the first time the student feels when listening: "That is of great interest to me! - at this very moment the long, enticing way of a documentalist starts, rich in advantages, in experiences, in dissappointments, which nobody knows but you yourselves. That means in consequence, not to strive methodically for the moment in regard to which Faust longing says: "Linger a while, you are so beautiful". As to the method we can therefore proceed in 3 different ways:

1) with pathos, viz. complaint, praise, heroic songs of praise, gilding, accentuation of culminating points in development, underlining of evident pluses or minuses are the most important elements.

2) with a irony viz. the lecturer reflects ironically on the deficiency as an effect arisen in the subject. Tension between longing and fulfillment is reflected; a deficiency is represented by irony as difficult to overcome and hardly avoidable. Or, a deficiency may be represented as cultivated laziness or mistakes may be unmasked in a wrongly programmed mental range or unpleasantness of life occurring everywhere in library and documentation service, and which I meet too - may be presented.

3) with a sense of humour, viz. the powerful palliative against a deficiency not yet overcome, against incomplete intelligence and lack of knowledge.

Therefore, it seems important to me to take into consideration - for any education needs lecturers - that an educational system for lecturers in library and documentation services should be organised at once. For a high standard of knowledge doesn't qualify a lecturer or a teacher: Most of the library and documentation schools find their lecturers in the same way, as one finds a $100.- note, by accident, unsystematically and most of them do not find even one. They mostly find cents, that means "in general: the teacher is not prepared for the library and documentation centre of tomorrow, but for the one of yesterday."[10]

The position of a professor in the university's scope or of a director of a museum, or of library or documentation centers does not guarantee the ability to teach or to act as lecturer. For not only the subject matter should be taught, but it should be shown also how to develop the subject and how to acquire a personal relation to it. Even the lecturers' work should be submitted to control, censorship, and examination. Qualification of the educators, lecturers, and places of education must be fixed. The authorization to educate should be limited to three years. A control of education courses and the determination of educational aspects should be introduced. A most specific educational strategy against inflexibility and inefficiency must be developed.[8] We must try to improve a great number of relations between the educational system and the educational environment. You see that the main problem is: how to arrange for an active educational system. Modernization of education also depends on providing the educational system with means for innovative processes and on creating an adequate modern administration. In fact: all desirable measures of performance are reducible to the user criteria of reliability and efficiency - and the management criteria of cost. Otherwise, it only represents a pretended school, a pretended education. The critical self-control to which students should be raised, should be set as an example and demonstrated by the lecturers who ought to have this ability.[2] A better control of the results in education should be introduced.

But an education for lecturers which is oriented towards application and practical work does not exist.

The greater is the uneasiness in special libraries or documentation centers, which employ scientists who are not at all or poorly educated, that means, uninitiated employees, who have ideas but no complete educational background. Consequently post-graduate seminars for lecturers - not only for information officers - should be regularly given in all countries of the world in order to lead out of ivory towers and to bring every routine training programme to life.

I am closing my lecture with the following old story, which will demonstrate, that this fear is very old and that hope and good wishes are justified.

"When I was on my travels, I met a huge haycock which had collapsed and lay across the road. The peasant, standing beside, called to me to help him to set up the cart. That touched my heart. We had two planks to hand. We forced it under the cart and lovered tp the best of our ability, the vehicle tottered, rose up and stood and we loaded the hay again. And the peasant said: "Nobody knows whether he can do something unless he has tried it and because you got in my way. The cart collapsed, in order that you may help me. What else have you learned if you haven't learned that everyone has his own way of duty? What would be a God who only would have one single way of service? He doesn't exist. There is a way to serve God by praying and there is a way to serve Him by good deeds for other people and there is another way to serve God by learning.

This story is not my own story, I found it in the book "Gog and Magog" written by Martin Buber.

It symbolizes our subject, for the theory of library and documentation service is actually not different from the assistance to the great number of "hay carts" that have collapsed. Even lecturers of information science should learn to see those "hay carts" and to draw the listerners' attention to them. We all should help, setting up the "hay carts" collapsing every day. We all should wish that it were only collapsed hay carts needing our help.

REFERENCES

1. GOETHE : Artemis-Gedenkausgabe, vol. 8, 1961. p. 104

2. GOETHE : Artemis-Gedenkausgabe, vol. 8, 1961. p. 607. Goethe's statement on the author, is also valid for the lecturer: "For the author rarely is unprejudiced judge of his own works."

3. KENT, A.: Information Sciences-Development of an interdisciplinary graduate education program at the University of Pittsburgh. (American Documentation Institute. 26th Annual Meeting, Chicago, Ill., Oct. 1963. p. 111-114).

VI Selection, Education and Training of Personnel

4.	GOETHE :	Artemis-Gedenkausgabe, vol. 14, 1964. p. 915.
5.	GOETHE :	Artemis-Gedenkausgabe, vol. 22, 1964. p. 651.
6.	NATHALIE SARRAUTE :	Cited by Ludwig Marcuse: Argumente und Rezepte. München 1967. p. 60.
7.	VINOKUR, M.M., et al.:	Request: "considerable encouragement of librarian activity for the improvement of the system of recommending bibliography of technology. Complex planning, investigation of theory and methodology, renovation of classification, critique of literature and summary and expansion of the best experiences." (Vinokur, M.M. and others: Technical bibliography and bibliographical information activities of the technical libraries. Sovetskaja bibliografija, No. 3, 1967. p. 91-131.)
8.	COOMBS, P.H.:	World Educational Crisis. A System Analysis. Oxford 1968.
9.	HOLM, B.E.:	How to manage your information. New York. Reinhold Book Corp. 1968. p. 292.
10.	COOMBS, P.H.:	Die Weltbildungskrise. Stuttgart 1969. p. 187 and the chapter: Teacher - a problem of quality and cost. p. 50-61.
11.	GRUNWALD, W.:	Asks "For the proposed system of education we are in lack of teachers. This objection reveals a serious difficulty. We are indeed in lack of teachers, especially of teachers for information science". (Der Bibliothekar und seine Ausbildung. Zeitschrift für Bibliothekswesen und Bibliographie 16, 1969. p. 168).
12.	MARTYN, J., and VICKERY, B.C.:	The complexity of the modelling of information system. (Journal of Documentation 26, 1970. p. 204-220).
13.	TAGLIACOZZO, R., ROSENBERG, L., KOCHEN, M.:	Access and recognition. From users' data to catalogue entries. (Journal of Documentation 26, 1970. p. 230-249).

GRADUATE EDUCATION NEEDS OF INFORMATION SPECIALISTS

Dorothy B. Lilley
Professor and Head
Department of Library Science
East Texas State University
Commerce, Texas, U.S.A.

SYNOPSIS

This study identifies the graduate education needed by the practicing information specialist: (1) subject reference and subject bibliographic control, including control of current information; (2) research in information problems; (3) information center administration, including cost analysis; and, (4) design, evaluation and use of information systems.

Introduction

Administrators in ten selected scientific information centers in the United States were interviewed in 1964 (5). The centers, representing eight different scientific and technical fields, were selected for their use of sophisticated, nonconventional information systems. Data were collected by means of a structured interview schedule. These data were supplemented by those from interviews with officers of professional organizations in scientific, technical and special library and information science fields, and by reading and consulting during the next five years while analyzing and interpreting the extensive data.

The purpose of the study was to find out (1) to what extent information centers were different from libraries and different from each other; (2) what activities information specialists performed or in which they were most frequently involved, for it was, in fact, found that several different types of professional personnel were involved in most of the information center activities; (3) what knowledge information specialists needed; and, (4) what part of the knowledge that information specialists needed required formal education.

Apparently by selecting outstanding centers in terms of their sophisticated information systems, and studying these in detail, trends were identified that have prevailed in the educational needs of practicing information specialists. Factors which affect the kind and extent of the formal education that information specialists need are: the environment in which the information specialist works, the activities in which he is involved and the areas of knowledge that he can learn on the job.

Environment

The environment in which the information specialist works is different from the university library environment. Three types of information centers that are different from libraries and different from each other were identified. The differences between the information centers and the libraries were attributable mainly to differences in input, i.e., personnel, materials and equipment use, and did affect a difference between the education needed by librarians and the education needed by information specialists. The differences in input between the libraries and the information centers were (1) several different types of professional personnel were employed by information centers: computer specialists, scientific and engineering personnel, information specialists, librarians and administrators, while in the libraries there were only librarians and administrators; (2) several types of documents different from those in libraries or no documents at all were input by the centers; and, (3) input of computer and other equipment was widespread in the information centers, but not, at that time, in libraries. The differences among the information centers were largely differences in distance from and relationship to the users, some being totally remote, others working with their users in a common endeavor and others acting as liaison officers from sources of information to the user, who in these cases, were research scientists. These differences did not appreciably affect the activities of the information specialist in the different types of centers, not the kind of education needed.

Although this study did not expressly investigate the differences between the output, i.e., services, of the libraries and information centers, it is known that there are differences in form, e.g., print-out from microform; method of serving the user, e.g., selling or giving the output rather than lending it; and, differences in relationship to the user, e.g., searching out and delivering information to the user rather than requiring that he select and retrieve it himself. Some of these differences are more a matter of degree than of types of services. One of the principal differences between libraries and several of the centers is that some centers operate for profit or to break-even. This is a trend that has been developed to a fine point in one center investigated in 1970 by a graduate student at East Texas State University (7). Using the same interview schedule which was used in the original study, he found that all of the services of the center are on a charge basis with a subscription deposit against which all services are charged.

Activities and Areas of Knowledge

The activities in which the information specialist is involved and the areas of knowledge which he needs to know are closely allied. Seventy-one activities and thirty-three related areas of knowledge were surveyed. In the majority of the centers, information specialists were found to be involved in fourteen of the activities and to need formal education in thirteen of the thirty-three related areas of knowledge. The information specialists were considered to be relatively more competent in the activities and areas of knowledge related to design and operation of their own information systems: coding index terms, developing terminology authority, information system design, transforming user questions into the terminology of the system, evaluating materials retrieved, coordinating elements of the information system, retrieval by logical operations and selective dissemination. They were less competent in activities and areas and of knowledge related to subject bibliography and reference, research in information problems and information center administration, including cost analysis. The former are the areas that have tended to define information science. Several engineering and technical schools have developed educational programs including these areas.

Graduate Education Needed

The principal contribution of this study is that it revealed (1) that the function of the information specialist is largely locating and transmitting subject information to the user; (2) that knowledge of subject bibliography and reference and bibliographic and subject matter control is the greatest felt need for the education of the practicing information specialist; and, (3) that therefore, the program of education in information sciences which does not include the study of guides to and sources of the litérature of a subject, including the available subject data banks stored on magnetic tape or disk, cannot be considered adequate education for the practicing, professional information specialist. This fact contributes to establishing the library school as the location for education for the information specialist. These findings have been obliquely confirmed by Klempner (4) who reported in 1968 that the information from several centers was isolated and poorly disseminated, implying that information specialists would benefit from knowing of other information besides that produced by their own centers, and by Taylor (3) in 1967, who, although not affiliated with a library school, recommended that library schools change their curricula to parallel the growth of the information sciences and that graduate study in the information sciences be imbedded in research. While library schools have traditionally accepted the responsibility for education in subject bibliography and bibliographic control, they have little tradition in research and evaluation testing the status quo and still less, until recently, in design of nonconventional information systems. The problem has been, as Muller (1) noted, that library education has been taught as "Bible-truth" while it should be taught as "an area of knowledge and theory that is subject to continuing reappraisal and renewal." The design of new information systems was a major break from the tradition. It has incited a new type of service and a new type of operation. This study found that administration of such services through information centers placed major emphasis on cost analysis. This is understandable since there are a number of centers which operate for profit. At the time of this study there were very few courses being offered in the administration of information centers. The content of such courses is a matter of current concern (8).

Conclusions

The principal areas of knowledge that require formal education are those related to subject reference and bibliography; research in information problems; information center administration, including cost analysis; and, information system design, evaluation and use.

This study, although based on data from scientific and technical information centers, predicted widespread application to other fields. Wider application is indeed the current prognosis, not only in information centers, but in libraries (2)(6).

References

1. Muller, Robert H. "The Research Mind in Library Education and Practice," Library Journal, 92, (6), March 15, 1967, pp. 1126-1129. p. 1128.

2. Myatt, Dewitt O. and Barclay, Donald A. Position Paper of Extra Library Information Services. Final Report. Conducted for the National Advisory Commission on Libraries Under Duke University Contract OEC 2-7-010105-1523. Washington, D.C. Science Communication, Inc. Dec. 15, 1967. ERIC ED 023 409.

3. Taylor, Robert S. Curriculum for the Information Sciences. Report No. 12: Final Report on Curriculum Development in the Information Sciences: Recommended Courses and Curricula. NSF-GE-2569. Bethlehem, Pennsylvania. Center for Information Sciences, Lehigh University, 1967. pp. 32-34.

4. Klempner, I.M. Diffusion of Abstracting and Indexing Services for Government Sponsored Research. Metuchen, New Jersey, Scarecrow Press, 1968.

5. Lilley, Dorothy B. "Graduate Education Needed by Information Specialists and the Extent to Which it is Offered in Accredited Library Schools." Unpublished D.L.S. dissertation, New York, N.Y., School of Library Service, Columbia University, 1969.

6. Lancaster, F.W. "User Education: The Next Major Thrust in Information Science?" Journal of Education for Librarianship. 11, (1), Spring, 1970. pp. 55-63. See footnote p. 62.

7. Lyde, James P. "A Comparison of Two Information Centers and a Research Library in Dallas, Texas." Unpublished paper, Commerce, Texas, East Texas State University, Dec. 12, 1970.

8. Borko, Harold. "Predicting Research Needs in Librarianship and Information Science Education." In: Proceedings of the American Society for Information Science, 7, 33rd Annual Meeting, Philadelphia, Pennsylvania, October 11-15, 1970. pp. 27-29.

Session Six - Discussions

SELECTION, EDUCATION AND TRAINING OF PERSONNEL

Chairman: Prof. J. Gross (Israel)

PROF. H. BORKO (US): Some of you may be familiar with a curriculum that was published in "Computer Sciences" by the Association of Computing Machinery. This had a great impact on developing a separate discipline of computer science in the US and throughout the world. The Association is now working on another curriculum for a degree called "Information Analysis and Systems Design". This will be published after it has gone through all the necessary reviews, which will be in six months to a year. I think it will also make a great impact and there is a need for discussion of it from our point of view so that the new curriculum, whatever it will be, will not be finalized before we have got our point of view across.

MRS. L. VILENTCHUK (Israel): I want to touch on the shortcomings of most training programs in information science and the reasons for these shortcomings as I see them. The training of information workers in most countries, including Israel, is like an upside-down pyramid with emphasis on theoretical subjects. What, in effect, are we doing? We are trying to produce research workers and theoreticians instead of preparing a cadre of useful and highly qualified technicians. When I say technicians, I don't necessarily mean non-graduate workers. In my terminology, technicians are all those who practice their profession as opposed to those engaged in research and development of new techniques within their professions. They may be non-graduates, or on a higher level, graduates with Bachelor's or Master's degrees.

University schools in information science are sprouting all over the world. The curricula of most of these schools either impart knowledge in library techniques only, or, going to the other extreme, emphasize purely theoretical subjects, such as mathematical theory of communication, operations research, etc. Even worse, in my opinion, they emphasize the study of computer technology as a subject in information science, disregarding the fact that an information scientist should be the user of library and computer services and not their operator.

The students are fed a lot of inapplicable knowledge, rather than being given the tools for the intelligent practice of the profession whose object is to serve as the liaison between the generator and the user of information.

Graduates are often unhappy and frustrated because their work after graduation does not measure up to the high expectations evoked by the printed curricula of the schools. Moreover, the schools mislead them into thinking that the attaining of this or that degree guarantees proficiency on the job, whereas proficiency on the job requires long years of experience. The result is that the intellectually honest, the gifted, and those possessing initiative and personality are tempted to leave the profession.

The other weakness in the training of the profession is the criteria by which applicants for training are selected. Information science is a service profession, therefore, the personal qualifications of its members is of paramount importance. Unfortunately, no research has been done yet on what personal qualifications are needed in a competent information scientist. Because of the newness of the profession there is probably not enough data on which to base such research. However, common sense and the experience of managers heading information units may provide some useful guidelines.

We need, primarily, all-around, good general practitioners, to borrow a term from medicine; people with a broad range of interests, a sound general education and specialized education in the particular discipline in which they will work. People with intuition and the ability to feel, think and live other people's feelings, thoughts and needs; a command of several languages, a balanced, not aggressive personality; the ability to remember facts they come across even through chance reading or personal encounters and associate them at the right moment with the problem at hand; people who pay meticulous attention to details, but who nevertheless discern the essence of a problem and, finally, people who are articulate in oral and written presentation. Very rare qualities indeed.

The best age is just over 30; at that age students are still young, but not youthful; the very young are generally aggressive, egoistic, self-centered, unable to understand other people's psychology, needs and viewpoints - all qualities which are very good in some professions (e.g. the military), but not suitable for any service profession.

One last point. There are professions which require from the outset a great amount of theoretical knowledge, such as, for instance, chemistry: the education of a chemist today is not complete if he has not acquired a Ph.D. degree. On the other hand, there are professions like engineering or our own profession and many others, which require years of practical experience before additional theoretical knowledge is digestable, useful and desirable.

I therefore believe that schools with a Ph.D. program in information science are premature. The elite of our profession, after years of experience on the job, will develop new techniques, do research and open new vistas for the profession with or without a Ph.D. degree. These activities are not to be engaged in from the school bench.

PROF. P.N. KAULA (India): India is a Federation of eighteen states where education is a state subject. Library education is the responsibility of the individual states, but the Central Ministry of Education lays down the national policy for the entire country. The University Grants Commission sets the norms like the University Grants Committee in the UK.

There are 34 university library schools in 14 states. In addition, we have two information science schools which offer specialized training in documentation and information sciences.

The equivalence of the degrees in library science in various countries is a problem and this is the right forum to discuss it. A Master's degree program in my country is a two-year, full-time professional course given after the completion of the first degree, while in many countries a Master's degree program is only one year after graduation. Degrees awarded by the British Library Association and by the London School of Librarianship and others are not recognized as an equivalent to a Master's degree in the US. Such incompatibilities create problems for students who would like to go to other countries for higher education. Standardization of degrees is necessary.

The following extract concerns the equivalence of professional degrees in Canada, the US and the UK. I quote an extract from, "Baffling variety: educational methods for librarianship in Great Britain, Canada and the United States", A. Roberta Rogers, Coll. and Res. Libs., 23(1), Jan. 1962, p. 45-50.

"In the summer of 1959 an informal committee consisting of J.C. Harrison, Head of the School of Librarianship, Manchester College of Technology; George Pitcher, Librarian of the Kumasi College of Technology, Ghana; Bertha Bassam, Director of the University of Toronto Library School; Lester Asheim, dean of the University of Chicago Graduate Library School; and Robert L. Gitler, executive secretary of the Library Education Division of ALA, met in Urbana under the chairmanship of Dr. Lancour.

"The committee recommended the following table of minimum requirements:

Canada	a)	Bachelor's degree from an approved institution.
	b)	B.L.S. degree from a library school accredited by the ALA and CLA.
U.S.	a)	Bachelor's degree from an approved institution.
	b)	M.S., M.A., M.L.S. (or similar degree) from a library school accredited by the ALA.
U.K.	a)	Bachelor's degree from an approved institution.
	b)	One year of study in one of the ten library schools approved by the Library Association.
	c)	Possession of the Associateship of the British Library Association.

"These proposals mark a real step forward in the matter of international library cooperation, although they do leave a few questions unanswered. For example, no mention is made of the recognition to be accorded the holder of a British university degree plus the FLA or the Academic Post Graduate Diploma in Librarianship of the University of London. By implication at least, these qualifications are ranked somewhere above the minimum, perhaps at the level of the Canadian M.L.S. or the old-style sixth year American master's degree".

Work has already been done on the formation of an International Association of Library Schools, and we have been in correspondence with IFLA. We have about 35 library schools who are interested in forming an International Association of Library Schools to discuss library education and training programs. We should take some decision on this issue.

The next point to be considered is the exchange of teachers in library science in various countries, provision of fellowships for foreign students in selected library schools, holding professional international seminars on library education, analysis of various courses in library schools, etc. I would like to invite your suggestions on these issues which are pressing problems for library education in every country.

PROF. V. SESSIONS (US): The professional development program at the City University is intended for people who are practising librarians and for information specialists who want to learn some of the newest techniques. However, the program was designed with a broader goal in mind. We were trying to look at what librarians and informations scientists in general have to know, not just what practitioners have to know.

We may, as we go on to the doctorate, have an intermediary step. It is possible that we will decide to have a broader discipline called Information Science, which will be different from Computer Science. I did have several discussions about the eventual Ph.D. degree in Computer Science with the Head of the new Computer Science

Program at the City College. It would be a travesty to give a Ph.D. in Computer Science to somebody who will only have the level of proficiency in Computer Science necessary for people who practice in libraries, information centers, data archives and such institutions. We feel that library or information science generally, and computer science, are quite different disciplines.

MR. H. SCHUR (UK): I will deal with four topics. First, why do we bother with education and training programs? As has been said before, there are quite a number of people who have no formal qualifications in information work or librarianship and yet are quite successful.

Second, if we have training programs, what are their objectives? Some of the dogmatic statements heard this morning stem from the fact that the objectives, in the view of the person who stated them, are very specific, whereas objectives which might exist in other environments are not taken into account.

Third, there are constraints. It is of no use just comparing programs in different countries because the conditions under which these programs had been developed are different. I am not trying to say that no two programs can be similar, but that we ought to know a little bit more when discussing programs, about the constraints and incentives under which these programs operate.

Fourth, what criteria exist for evaluating programs? There ought to be some sort of feedback on whether these programs are really useful.

I feel that there are really two major objectives. First, there are educational objectives which, so far as I have seen in the literature on educational training programs in this field, seem to have been completely forgotten. The objective of an educational program is educational. This may sound a platitude, but it is absolutely essential. By educational I mean Bloom's reasoning on education objectives. You will see that this takes into account some points like knowledge, which includes the recall of facts and generalizations of methods and processes, patterns and structures. Then one goes to the much more difficult level of comprehension, publication, application, analysis, synthesis, and finally evaluation which includes judgements.

I am afraid that many programs seem to concentrate largely on the knowledge part because it is very easy to test this. It is very difficult to test evaluation judgement, because people have got their biases and it is difficult to agree that somebody has made a good case if one holds a diametrically opposed point of view. However, it is the function of a teacher to be able to do this. So knowledge is the thing which tends to be stressed and people are walking encyclopedias.

Secondly, there are professional or utilitarian objectives. Obviously people do not attend courses just for the fun of it. There are four major types of utilitarian or professional objectives. There are those of the individual concerned, there are academic ones, those of the profession, and finally the national objectives, which are very pronounced in a country like Israel.

It seems to me that there are essentially four categories of work. First there is the production and operation category. This requires people who know something about the subject they are working in - the subject specialist. These are the practitioners. One also requires system operators. Very many librarians are just of that type. They know how to run the library.

Increasingly important, I think, are people who are capable of doing development. One knows of libraries which become fossilized. Development has to be continuous.

Applied research, which does not deal with the topics which bring prestige to the individual, are the topics which are essential to the country. Unless they are dealt with, the work won't get done.

Background research, which is almost pure research, appears to be essentially academic. I do no think one can really prescribe what sort of thing ought to be researched because this is something which comes from the individual himself. All that can be done is to state the sort of problem which exists, at a national level, and hope that some people may get interested. It is no use coming around and saying "you will work on this." Research is like the work of an artist. An artist does not normally produce works which have been dictated to him.

What are the criteria which ought to be considered? First, effectiveness; second, efficiency; and third, the benefit-to-cost ratio. Educational effectiveness: how effective is the course? Professional effectiveness: is the output of the program useful? It it actually used? Do the people in fact get jobs? Efficiency is essentially the graduate-to-cost ratio. It is certainly not the output-to-input ratio. One could easily pass everyone and be 100% efficient, but quite ineffective. Benefit-to-cost is extremely difficult to assess. There is a lot of discussion in educational circles about the benefits of education in general.

PROF. M. KOCHEN (US): To Prof. Borko. Tony Etinger of Havard has questioned whether it is desirable to have a separate computer science department; the same thing could be asked about information science. I wonder what your views on this question are?

PROF. BORKO: To Prof. Kochen. This depends upon your emphasis. We could emphasize it in our library schools; for example, UCLA gives a special degree in information science in their library school, and grants two separate degrees: Master in Library Science and an M.Sc. in Information Science. UCLA will give degrees in information science within the Computer Science Department and the School of Management. My own view is that there is enough of a common body of knowledge in information science and particularly in the systems analysis aspects to deserve a separate degree program.

PROF. SESSIONS: To Prof. Borko. Logic does not always determine where a program is located. Sometimes it is a matter of political or administrative necessity. Let us be honest about why some programs are called by certain names and why they are OK where they are. There are some theoretical reasons, but there are also some pragmatic reasons.

DR. J. ROTHSCHILD (Israel): To Prof. Borko. What are the undergraduate requirements for the different programs? Can one enter one of the computer programs or information science programs with a BA in the humanities?

PROF. BORKO: To Dr. Rothschild. This is a very important point. At the moment we see the level of technician as a terminal type of degree and the student will probably come from an equivalent of a high school education, maybe with some work experience and possibly some technical training. In Germany, this is the equivalent of a degree from the Deutsche Gesellschaft fuer Dokumentation, an institutional type of degree.

For the level of Information Specialist, requiring a Masters Degree, there is an area of specialization. The person who wants to work in sciences should have a background in sciences. If he wants to work in education or linguistics, he should certainly have a knowledge of more than one language; if he will emphasize management he would need a Bachelor's degree in management. He would also need, and this is a personal opinion, a fairly strong background in mathematics. These are the requirements of my school. For the doctoral level, he will have to do research work and this would have to be interdisciplinary.

DR. M. KESSLER (US): To Prof. Borko. This is not a question, but a request. Those in the profession who occupy themselves with matters of education should consider that the field is new and, therefore, by necessity, practitioners in the field have not gone through any formal education guided by our present understanding of what education ought to be. It is fair to say that this condition will prevail throughout the next 5, 10, or 20 years.

I would suggest that the people who give thought to the matter of education also concern themselves with the problems of education of the present practitioners whose circumstances are quite different from those of a student who attends classes from 9 to 3. They work all day and have to take evening courses. I think that considering the state of our technology this educational problem is of great importance and I would be very happy if the educationalists in this field would give some thought to it by organizing courses or even writing textbooks.

PROF. BORKO: To Dr. Kessler. I am glad to say, unequivocally, that people in all countries of the world are giving a great deal of attention to this and indeed most of the development has come first from in-house service courses, later from special technical schools and only then have they been adopted by the universities. The need to retrain and to continue the education has grown out of the technical level and there is still a major emphasis on this in all the countries of the world.

PROF. H. ARNTZ (Germany): I would like to add something from the point of view of the International Federation for Documentation (FID). We are in the process of setting up a new program and new statutes and we deliberated in Moscow whether to retain the word "documentation" in the name of the organization or to change it to something new. As this proved to be impossible, we decided to stick to the old name International Federation for Documentation with the understanding that "documentation" in the name meant Informatica, Dokumentation, Documentation, and Information Science, singular, but at the same time we agreed that our "documentation" was part of a bigger science called the Information Sciences, plural. Other sciences either belong to this group completely, e.g., library or archive science, or in part, applied linguistics, theory of communication, computer science (informatics), and so on.

PROF. GROSS: I think it is a tour de force to get a form of Esperanto into the name of an organization; you seem to have succeeded in that!

DR. ROTHSCHILD: I am glad to endorse at least two of the points which have been made. First, the importance of selection of trainees as well as of teachers. I am extremely glad that Mrs. Vilentchuk pointed out the need for general background, not only in the limited field of science or technology, and I am even ready to agree to the point she made concerning the desirable age. I like to have in my classes at least some students who have accumulated experience in life, served in education, etc.

The second point is service versus research. Our universities here are rather conservative. Only next year I hope to get the right to grant Master's degrees. Even today members of the University Senate doubt if ours is a scientific profession, and it is especially difficult to recruit science graduates.

Before the opening ceremony I had a talk with Dr. Penland and he asked me how many librarians from public libraries are at this Conference. I had to answer - very few. But public librarians also must have basic knowledge about all that is dealt with at this Conference. Concerning the recruiting of science graduates, I should like to mention that Mr. Lebowitz is more optimistic in this respect than I. He intends to interest science graduates at the Technion in joining a Master's program in library science.

In the States there is a five year program without any relation between the Bachelor's program and the one year course in librarianship or information science. I think it would be desirable to shift some of the basic subjects to the undergraduate level without giving Bachelor's degrees in information science and librarianship.

PROF. ARNTZ: We sometimes simplify the problem if we do not think of the persons whom we are going to train. What we really want are persons able to serve user profiles. This means they must have experience in the fields in which they will serve their users.

On the other hand, Mrs. Vilentchuk said that it is difficult to recruit people over thirty. This is absolutely true as I have found many times, but we want the best chemist to serve a chemical user profile. This means that we take him away from chemical work where he could make discoveries and become famous. We want the best medical doctor for medical profiles. We take him away from his medical work and so prevent him from becoming a famous surgeon. Only the best people in the filed will make the best information officers.

So we must give a lot of incentives to those whom we want as scientific information officers. Questions of pay, social status, general recognition of information work, and so on are involved. And this in addition to all the problems of curricula.

MR. R.B. ZAAIMAN (South Africa): To Prof. Arntz. In our experience, information officers are normally not people who are extremely successful in research or who find research very interesting. It is not those people who become good information officers. Good information officers normally start off with a highly idealistic view of science and research and become disillusioned. They then turn to information work in which they find more intellectual stimulation.

I think research institutes which have funds available can make a study of the psychology of the successful information officer. At the moment one has to find these people by trial and error. There is no way of actually selecting them according to psychological tests and one only learns by experience which people are likely to be successful and which are not.

DR. J.E. BROWN (Canada): I would like to confirm what Mr. Zaaiman just said. In Canada during the past six years we have operated an in-training program sponsored by the National Science Library and a scholarship program, in an effort to encourage scientists to come into the field of science information. We have attracted some very good people, but on the whole the applicants have tended to be people who are not doing too well as scientists and have decided that they have to get into another field.

As a result of developments in Canada, we are now undertaking a very detailed study of what type of training information scientists require. We have a team headed by a professor from Dalhousie University which is trying to find out what the training needs are, taking a look at the situation as it now exists in Canada, evaluating whether it is adequate or not, and if not, what should be done about it. I suspect that the needs in Canada are no different than in other countries and the team's report, which will be out within the next three or four months, will be of interest to other countries too.

DR. A. COCKX (*Belgium*): This discussion has emphasized the difference between the Anglo-Saxon approach and the European approach. The American approach over-emphasizes the information scientist as against the subject specialist. I cannot accept Prof. Borko's view that a Ph.D. in information science is a qualification for work in a chemical information center.

Dr. Gezelius has mentioned that libraries are creations of the nineteenth century and documentation centers are the creations of the first half of this century. When we enter the second half of this century, the dialogue in information transfer is only possible when there are top qualified people on both ends. If you do not have this dialogue your information center is a complete failure.

PROF. GROSS: I cannot think of the future of information as a mechanism without highly trained people at the Ph.D. level.

PROF. BORKO: *To Dr. Cockx.* I do not think there is any disagreement at all. What I was trying to describe are the different needs within the broad field of information science, and the different levels of training necessary to meet these needs.

I agree completely that for the service oriented degree there is no need for the Ph.D. It is necessary for a research and teaching career to advance the state of the art in communication, in information analysis and dissemination. Whether the same person can do both, whether he will progress by experience from the service degree and then come back to the university because he has a real research problem and wants to be able to pursue it, is a different matter. We should not make the mistake of trying to force all education into a single mould. The field seems to be developing into various types of orientations and various levels of competence.

Not everyone is going to go into management and information centers. It is for the university to decide whether all those going into librarianship or information work should have courses in cost accounting, in mathematics, in operations research. There are different levels of specialization. It would be a mistake to say that a particular curriculum is the only curriculum.

MR. ZAAIMAN: *To Mrs. Sessions.* You said in your paper that the City University requires those students who do costing to have quite a good background in mathematics and statistics. That is very true if one works at a doctoral level, but I think that any librarian who wants to do costing would be very lucky if he had the very specialized training in costing and statistics required. In most instances the services of specialists in these fields could be called on.

PROF. SESSIONS: We feel that a certain basic elemental understanding of costing and management is essential to almost everybody at some point in his career. Perhaps not for the first degree, but certainly later. I have been asked to chair a seminar of the Library Association on new developments in librarianship, for the City University next spring. We have 276 libraries at the university. We asked ourselves, once they know about new developments, how will they be able to apply them in their own libraries and what will the effect be on the university library system? So we decided to make this into a 2 year program. The first year is devoted to matters that we feel almost everyone should know, but is not aware of, and therefore, not really capable of making the kinds of decisions which should be made on the low level of the hierarchial structure.

PROF. ARNTZ: We in Europe were unhappy when the Americans changed documentation to information science because the word scientist here takes on a different meaning from thar which it normally has. If information science is a science, and we are convinced that it is, then we need information scientists who are scientists in the field of information science. The man of whom I previously spoke - the best chemist who is to serve user profiles in chemistry, the best medical doctor who has to do this in medicine - he, of course, just as Dr. Cockx said, must have studied chemistry or medicine and in addition, after practical experience he must have studied information science, but this is not his subject field. The man who runs a big information or documentation center, in a chemical plant, asked about his profession will always answer, "I am a chemist", but, of course, he is also an information man.

So we in Europe and in the international field distinguish between the real scientist in information science whom, as I said, we badly need, and the man who serves user profiles. He has studied a particular subject but afterwards he certainly has studied information science as well.

MISS M. ATIR (Israel): We forgot one very important thing and that is the public. What is the image of our profession from the point of view of the people we serve? I studied at Columbia University in the States and when I started working in the New York Public Library people asked me - what is a librarian? Do you need to study to become a librarian? I hear the same question in Israel. The public does not know what our profession is. How can a student want to become a librarian if he does not know what it is?

PROF. GROSS: I can talk as one of the users. I am not a librarian. I use libraries. Research people learned very quickly that they waste an awful lot of effort in duplication if their interaction with the information source is not adequate.

MRS. E. EHRLICH-DEVRIES (Israel): The overlapping disciplines demand an enormous amount of special knowledge from a librarian and I wonder whether in a special librarian's training it would not be possible to have two years of general education in addition to the librarianship subjects and in the third year specialization in the field in which the student would later like to work. A degree in civil engineering for example, is not much help to somebody who will later work as a librarian in medicine. If, on the other hand, in the third or fourth year the student learns a special discipline, it might afterwards give him an enormous advantage in finding work.

PROF. BORKO: To Mrs. Ehrlich. This suggestion is being followed now by a number of institutions in many countries. Let me speak for a moment about my own institution. We started with a one year Master's degree, and immediately recognized the problem that you raised, the need for specialization. A medical librarian needs different training at the specialization level than a military or a rare book librarian and so a second year has been added, which is a sixth year to the whole program.

It now includes a four year Bachelor's degree, which is the basic first degree; then the one year Master's degree in librarianship; and then the sixth year for a certificate of specialization. Now, UCLA felt that we should make it more general so we are beginning this year a 2-year Master's program. The first year of our Master's program will deal with general subjects in librarianship: basic work in classification, bibliography, reference and management; and the second year, after the student already knows what he wants to do upon graduation, with specialization. Many other schools are considering this same approach.

PROF. GROSS: To Prof. Borko. It comes to mind that with 6 years of training you are getting pretty close to a medical degree and I wonder why an internship in an appropriate library is not included.

PROF. BORKO: To Prof. Gross. Our 2-year degree program does indeed include a summer internship program in the second year. The question is whether it is worthwhile, cost effective, to go to the extra year? Is there an increment in salary commensurate with the cost of that training? At the moment the thinking at our University is that jobs are difficult to get and that a person with an extra year training would be more readily selected for a job, other things being equal, and we hope that this person will start at a higher salary.

PROF. SESSIONS: There is, of course, a great deal to be said for the idea that there should be specialized training in medical librarianship, engineering librarianship, etc. I have had a great deal of legal background, so I obviously took a course in legal bibliography and in the social sciences. However, in the US there is a great deal of moving around from one type of library to another. I think that we are doing a disservice to librarians by limiting them to one type of library. If librarianship and information science is a discipline in itself, then there are basic techniques in information science which can be applied no matter what the subject field is.

MR. SCHUR: I have in fact put this into my paper when talking about professional objectives. In designing a program consideration must be given to the career expectations of the graduates, which must not be worse than expectations of graduates in related fields, in terms of financial reward, prestige, interest in the subject, etc.

We must avoid developing programs which lead to dead-end jobs, however great the demand may be for such personnel. On the contrary, we should aim at providing an education for flexibility. The education for flexibility will not only increase the employment possibilities of the individual, but also by exposing him in the course to completely new topics, will broaden his outlook and hopefully increase his interest in the subject.

The point I am trying to make is that these special purpose courses for law librarians, for medical librarians, and others, which in fact do exist in some schools in the US, are highly undesirable as a first basic program. I am stressing this as a first program. One does not train a toothpaste organic chemist, but expects him to take a general chemistry degree, which includes organic, inorganic and physical chemistry, and only then allows him to specialize. We ought to be doing the same. People need a broad background.

MRS. VILENTCHUK: The status and image of our profession are at present not good. But I have many colleagues who's personal status is excellent. Why? Because status is something which you don't get by resolutions or with a high salary, but through your personal qualifications and the usefulness of your work to the community. And here again I come back to the Ph.D. problem. Ph.D. training is probably necessary - but for whom - for us, for the professionals, to advance our profession. It is not necessary, however, at the moment, for our users. For our users what is necessary is the technical level - people who can give really useful service.

The Ph.D. programs may be necessary later, and I come always back to the example of Lomonosoff. Somebody who really has something in himself and has something to contribute cannot be held down through any external circumstances. Lomonosoff was a Russian peasant who lived at the beginning of the 19th century when there still was serfdom in Russia. He walked by foot from Siberia to St. Petersburg motivated by his thirst for learning. Born a serf, he ended up as the first President of the Academy of Sciences of Tzarist Russia. Why? Because he had it in him.

We do need people who will advance the profession and they will probably need Ph.D.'s, but if they have it in them they will do it without a Ph.D. too. But we must not build an upside-down pyramid. We must first improve the technicians, the engineers of our profession who will be immediately useful to the public and will help to improve our image and our status.

MR. D. ELAZAR (Israel): I would like to make a comment regarding specialization. I think we have not taken into consideration the overlapping in the sciences. You can't take an information officer and say - you are going to be an information officer in chemistry - without understanding something about physics, mathematics and possibly geology and engineering. From my own experience I can say that the fact that I was able to work in a general science library helped me later immensely with specialized information work.

MRS. F. CESTAC (France): I would like to comment on Mr. Reichardt's paper. We train students to be information specialists and we have a two year program in information science. The first year is devoted to library oriented subjects and to computer work. We insist on a dialogue between the trainer and the trainee and apply a method of group dynamics.

We require from the students special competence in a specific field, such as management or systems analysis. In the second year we teach system analysis and management, but we also have special training in semiology which nobody here mentioned though it is very important.

I have no scientific background but rather a literary or artistic one, and I have noticed that often documentalists have no artistic feeling at all and know nothing about graphical layout. This is an important point and we have had a great demand for it in France. We have also started courses in oral communications.

PROF. KAULA: In library education much depends on what the library schools and teachers are able to do. A good curriculum with bad teachers cannot produce good librarians. We should have good teachers with good backgrounds in the subjects they teach. Normally when we recruit teachers we don't ask whether they have work experience in a particular library or a specific documentation field. This should, however, be considered when the teachers are being recruited.

I am pleased to hear that UCLA now has a two-year Master degree program. In India, we have had two-year Master degree programs since 1948.

DR. S. SCHWARZ (Sweden): As you know, Martin Luther said that the way to hell is paved with good intentions. At various Conferences, not the least so in futurology in which I am working now, I hear these good intentions have never materialized in really good propositions.

Now the question of usefulness is quite interesting. We say, "Let us be more useful". But I am not sure that esteem is related to usefulness because for instance, a garbage collector is one of the most useful persons in society and yet the status of the job is not that high.

I believe that the relation between information and the structure of acience and of scientific growth has not been really considered. If information could go into science policy, which can be easily done, by a science index method or other similar methods, we could show how useful we are and extend the field in a direction which can then be related to esteem.

SESSION SEVEN

publishing and reprography

DEVELOPMENT OF SPECIAL METHODS TO PRODUCE MICROFICHE ECONOMICALLY IN SMALL QUANTITIES.

Georg Thiele
Arbeitsstelle für Reprotechnik im Institut für Dokumentationswesen
Frankfurt / Main

SYNOPSIS

The increasing flood of literature in science and technology makes it necessary to find new ways for storing records and information. Large documentation centres have been using microfiche successfully for many years. The technical possibilities of using the advantages of microfiche for storing information even in limited quantities are explained.

Mit steigenden Veröffentlichungszahlen besonders im wissenschaftlich-technischen Bereich ist das Ende des Zeitabschnittes gekommen in dem Papier die einzige Materie zur Fixierung und Übermittlung von wissenschaftlichen Mitteilungen war. Neue Speichermedien werden gesucht, für die zwei Voraussetzungen wichtig sind. Bei einer Speicherkapazität, die erheblich höher sein muß, als die des üblichen Druckes muß ein schneller Zugriff gewahrt bleiben, oder verbessert werden. Die Effektivität einer Veröffentlichung darf jedoch durch eine Veränderung der Publikationsform weder eingeschränkt, noch erheblich verteuert werden. Normale Druckschrift oder Schreibmaschinenschrift läßt sich auf Papier höchstens um 50% verkleinern um noch lesbar, und was auch wichtig ist, kopierfähig zu bleiben. Magnetspeicher und Mikrofilm bieten sich an, da ihre Kapazität erheblich höher ist, wobei der Magnetspeicher zwar die geringere Kapazität, aber den schnelleren Zugriff bietet, der Mikrofilm dagegen die höhere Speicherkapazität. Es hat sich in der Dokumentation bisher als sinnvoll erwiesen, in Magnetspeichern keine Gesamtinformation zu speichern sondern hier Auszüge in Form von Schlagwörtern, Daten und Literaturangaben festzulegen, während es der Mikroverfilmung vorbehalten ist, Gesamtinformationen, seien es Texte oder Bilder, die als Information in der originalen Beschaffenheit erhalten bleiben, zu speichern. Die vorliegende Information wird verkleinert und in analoger Form archiviert. Man kann in diesem sachlichen Zusammenhang deshalb auch von Analogspeichern sprechen.

Besonders eingeführt hat sich hierbei das Microfiche in seinen international genormten Abmessungen von 105 x 148mm, mit einem lesbaren Kopf und fünf Reihen von je 12 Aufnahmen mit einem Verkleinerungsfaktor von ca. 20. Ein solcher Microfiche hat in der ganz normalen

Ausführung eine Speicherkapazität von mindestens 200 000 Zeichen. Stärkere optische Verkleinerungen sind heute bei anderen Verfilmungsarten bereits in der Anwendung und könnten auch die Kapazität des Microfiche noch erheblich erhöhen. Die derzeit internationale Form des Microfiche, die auch einer ISO-Empfehlung entspricht, ist aber bereits so weit verbreitet, daß man erwarten kann, allerorts Benutzungsgeräte, d.h. Lesegeräte oder Rückvergrößerungsgeräte, vorzufinden. Beispielsweise in den Vereinigten Staaten erfolgt die Informationsverteilung in Form dieser Microfiche durch den National Technical Information Service, wobei ein Tagesversand von 4 000 Microfiche als normal angegeben wird. Es ist naheliegend, daß auch kleinere Informationszentralen sich für das Microfiche als Informationsträger entscheiden möchten und zwar nicht nur weil die Lesegeräte an vielen Stellen bereits vorhanden sind, sondern auch weil es möglich ist, kleinere Microfiche - Serien in die großen Sammlungen zu integrieren. Diese Absicht scheiterte bis jetzt aber daran, daß die auf dem Markt befindlichen Microfiche - Aufnahmegeräte mit ihren sehr hohen Preisen für kleinere Dienststellen völlig unrentabel sind. Diese Geräte liegen in einer Preisklasse zwischen$20 und 35 000,-. Unsere Überlegung gingen daher davon aus, ein Gerätesystem zu schaffen, das stufenmäßig der Menge der herzustellenden Microfiche angepaßt werden kann.

Zunächst wurde sichergestellt, daß 16mm Mikrofilm - Aufnahmegeräte und zwar Schrittschaltkameras auf dem Markt sind, die genau dem international genormten Microfiche - Schritt von 11,75mm entsprechen. Zusätzlich hat dann ein Hersteller für Mikrofilmgeräte, die Mikrofilm GmbH, eine Tochtergesellschaft von Agfa - Gevaert, für ihre 35mm-Mikrofilm Schrittkameras, die in vielen Bibliotheken und Dokumentationsstellen stehen, eine auswechselbare Kassette für 16mm-Film mit einem festen Schritt von 11,75mm geschaffen. Hierdurch ist es interessierten Dienststellen möglich bei geringen finanziellem Aufwand Streifenfilme im Microfiche - Schritt herzustellen. Die Firma Kalle produzierte nach unseren Angaben einen Montage - Tisch, in dem der Film auf die Länge von 12 Aufnahmen geschnitten und festgehalten werden kann. Eine von der gleichen Firma hergestellt Acetat-Haftfolie im Format von 10,5 x 14,8cm, bei der die Schutzschicht streifenweise entsprechend den Zeilen eines Microfiche abgezogen werden kann wird dann auf den Montage - Tisch gedrückt und nimmt die Filmstreifen auf. Der noch fehlende Titel in der ersten Zeile des Microfiche,kann von einer Schreibmaschinenvorlage durch eine Schnellkopie mit Diffusionsentwicklung oder mit Hilfe anderer reprographischer Verfahren hergestellt und auf den Kopf der Haftfolie aufgetragen werden. Von dem so fertiggestellten Master für Microfiche können nun im bekannten Diazo - Kopierverfahren beliebig viel Microfiche kopiert werden. Eine solche Herstellung von Microfiche entspricht der internationalen Normung, da hier ausdrücklich festgelegt ist, daß auch die Kopien von Montagen, wenn sie der Norm entsprechen, und in der Qualität nicht von anderen Microfichekopien abweichen, als Microfiche akzeptiert werden.

Diese Aufteilung technischen Arbeitsganges hat den Vorteil, daß man bei geringem Bedarf mit der Beschaffung eines 16mm Mikrofilm - Aufnahmegerätes für ca. 1 500 Dollar bereits im eigenen Büro den Film herstellen kann, ohne Unterlagen aus dem Hause geben zu müßen. Der belichtete Film kann im einfachsten Fall an eine Zentrale gegeben werden, die diesen Film entwickelt, den Microfiche montiert und auch die Kopien herstellt. In Deutschland ist die Arbeitsstelle für Reprotechnik eine solche Zentrale, die derartige Arbeiten für andere Institute durchführt, die aber auch neben einem Dupliziergerät einen Rückvergrößerungsautomaten zur Verfügung hat, mit dem sie bei Bedarf auch Hardcopies von den bei ihr deponierten Microfiche herstellen kann. Die einzelne Dienststelle, die bisher nur den Film aufgenommen hat, kann dann bei steigendem Bedarf, durch Anschaffung der weiteren Geräte die Herstellung von Microfiche immer weiter, stufenweise, in eigene Tätigkeit übernehmen, ohne das vorher gekaufte Geräte überflüssig werden. Weitere Vorteile haben sich bei dieser Form der Microfiche - Herstellung ergeben, da beispielsweise Microfiche, die nicht abgeschlossene Berichte enthalten, durch zusätzliche Montage neuer Seiten auf den neuesten Stand gebracht werden können. Bei diesen Ergänzungen kann jeweils ein Datum mit verfilmt werden, um so den jeweiligen Stand der Arbeiten kontrollieren zu können. Ein weiterer Vorteil besteht darin, daß auch Halbtonbilder besser als bisher im Microfiche wiedergegeben werden können, denn es ist nun möglich, eine Serie Halbtonbilder auf entsprechendem normalen 16mm Kinofilm aufzunehmen und in den Microfiche einzufügen. Dadurch verschwindet der bisher sehr große Qualitätsunterschied zwischen der Schrift und der Bildwidergabe.

In Deutschland wird dieses Verfahren zur Herstellung von Microfiche, weil es problemlos und relativ billig ist, in steigendem Maße, besonders bei Instituten und Dokumentationsstellen mit einem kleineren Mengenbedarf benutzt. Es ist selbstverständlich, daß Dokumentationszentralen mit einem großen Durchgang von Reports auch in Deutschland auf automatische Aufnahmegeräte zurückgreifen, für die sich dann ja auch eine Rentabilität ergibt. Die Verbreitung des Microfiche wird noch weiter zunehmen, da bereits die ersten Microfiche - Farbfilme produziert sind, was besonders für die Medizin sehr interessant ist und da eine spätere automatische Sortiermöglichkeit bei Microfiche aller Herstellungsverfahren durch nachträglich angebrachte Code - Leisten in Zukunft möglich werden wird.

DEVELOPMENT OF SPECIAL METHODS TO PRODUCE MICROFICHE ECONOMICALLY IN SMALL QUANTITIES

Georg Thiele
Translated by the author from the German original

We note that the number of scientific and technical publications is increasing. The period during which printed paper was the only material for transferring and storing scientific knowledge is now coming to an end. Normally printed or typescript characters can be reduced up to 50%. But this is the limit if the characters are to be legible and - this is also important - it should be possible to make copies from them. It has become necessary now to look for new media in which information can be stored. These materials must meet two essential requirements.

1. The storage capacity must be much higher than that of printed matter.

2. Rapid and comparatively easy access to the stored information must be ensured.

Besides new media for storing, information should ideally not be less convenient than printed paper, and should also not be very much more expensive.

Two new media for the storage of information meet these requirements: Magnetic memory and microfilm.

The storage capacity of both media is much higher than that of printed paper. Comparing these two materials, the storage capacity of microfilm is higher than that of magnetic tape. On the other hand, the access to data stored in a magnetic memory is much more rapid. Therefore, in the field of documentation it has been proved useful to store in magnetic memories (tapes or discs) only such data which is relevant for retrieval purposes (keywords, descriptors, bibliographic data, etc.) and to store total information (reports, papers, tables, diagrams) on microfilm. In using this method the original paper is optically reduced and stored in analogous form. In this connection I would like to add that microfilm is an analogous memory.

Microfiche in the internationally standardized size of 105 x 148 mm is widely used. It is built up of an eye-readable headline and 5 lines with 12 pages on each. The 12 pages in each line may be 12 single pictures or 6 pictures with double-frames. The pictures have a reduction ratio of 1:20. The capacity of a standard microfiche is in the range of at least 200,000 characters. In other microfilm systems reduction factors higher than 1:20 are used, which makes it possible to increase the storage capacity of microfiche. The internationally accepted form of microfiche, however, which also corresponds to an ISO-Recommendation is meanwhile being widely used. It appears that the necessary equipment, "readers" and"read printers" for the internationally accepted size of the microfiche are available everywhere.

For example in the U.S. the National Technical Information Service is distributing information in the form of microfiche. On the average 4,000 microfiches are mailed by the National Technical Information Service everyday. It is obvious that other smaller information centers would also like to introduce microfiche for several reasons, as it is also possible to merge smaller series of microfiche into larger collections. In many cases the introduction of microfiche is prevented by the fact that the available micro-fiche cameras are very expensive. The prices vary between 20,000 and 35,000 U.S. dollars. In many cases this equipment cannot be used economically in smaller information centers. It was our intention to develop equipment for the production of microfiche, which can be adapted step by step to the number of microfiche to be produced.

At first it had to be ensured that 16 mm Microfilm cameras, flow cameras and planetary cameras were available which could take pictures in steps of 11.75 mm. This correspond exactly to the internationally standardized Microfiche step.

One manufacturer of microfilm equipment, Microfilm GmbH, developed an inter-changeable magazine for 16 mm film which has a fixed step of 11.75 mm. This magazine has been designed for Microfiche's 35 mm planetary cameras, which are in use in many libraries and information centers. So it is possible to produce microfilm 16 mm with microfiche-step. According to our specifications the firm of Kalle has developed a special mounter-table on which the film is cut into 12 exposure strips and the strips are then fixed in the mounter. The same firm has developed an acetate self-adhesive transparent film in fiche size, i.e. 10.5 x 14.8 cm. The fiche film is then pressed on to the mounting table and the film strips adhere to the transparent fiche film. The title of the microfiche can be copied from a typescript original using a Diffusion Transfer process or by other reprographic methods. This film strip with the title is then also mounted on the self-adhesive film. From this master a microfiche can now be produced, by using a diazo copying method. Microfiche made in the way just described is in accordance with international standards. It is said that mounted microfiche duplicates will be accepted as valid microfiche as long as they conform to the relevant international standards, and if they are of the same quality as other microfiche-copies.

The splitting up of the technical process for producing microfiche in sub-processes offers great advantages. If microfiche are only produced on a small scale, a 16 mm camera is sufficient. The price of such a camera is around $1,500. Microfilm can be produced at the office and there is no need to send out the originals. In the simplest case, the microfilm is sent after exposure to a service bureau which develops the microfilm, mounts the microfiche master and produces the copies.

In the Federal Republic of Germany, such services are run by the Arbeitsstelle für Reprotechnik at Frankfurt. In addition, there is also an automatic enlarger which produces readable hard copies of the microfiche.

But if microfiche is later to be produced on a larger scale, the office can purchase the additional equipment which is needed to undertake all the production processes on its own premises without having to discard the already available equipment. There is another advantage in utilizing this way of producing microfiche. For instance, a microfiche of a report, which has not been finished can be brought up to date by adding pages. In this case, the date can also be put on the microfiche so that the progress made in the work can be checked at any time. A further advantage is that in the most widely used negative microfiche, the continuous tone reproductions also appear in negative form. Now these pictures can be produced in a better way. It is now possible to mount 16 mm diapositives in the master microfiche. So there is no longer the difference in quality between the reproduction of text and pictures.

In Germany, this method of producing microfiche is being used more and more, because it is easy to operate and comparatively inexpensive. It is especially widely used in research institutes and information centers which are using microfiche on a relatively smaller scale. It is obvious that information centers which handle reports on a large scale will use automatic equipment which will then be more economic. There is no doubt that microfiche will be used more and more as recently it was shown that colour-microfiche could be also produced. This may be of special importance in the field of medicine. In Germany the first periodicals in the area of natural sciences will be published in the form of microfiche. For other purposes special code-strips are attached to the microfiche duplicate. This will also have the effect of increasing the use of microfilm.

REPROGRAPHY'S CONTRIBUTION TO INFORMATION SERVICES IN DEVELOPING COUNTRIES

E. Offenbacher

Head, Reprographic Service
The Jewish National and University Library, Jerusalem / Israel

SYNOPSIS

Conclusions about reprography in information services were reached in industrialized countries. They are not necessarily true for young nations. Differences are discussed concerning gathering of information and its dissemination. Economic considerations are stressed. Local technical and organizational limitations are taken into account. Training abroad is recommended.

Reprography, Information Services and Developing Countries are in themselves terms extensively considered in the literature. Much practical experience has also been gained concerning reprography's contribution to information services. For obvious reasons, theoretical as well as practical, conclusions have been reached in this field in the leading countries which happen to be the highly industrialized ones. No direct conclusion can necessarily be drawn from their experience for the benefit of the young nations. The foremost among the reasons would be the considerable difference in economic development and resulting possibilities. Organizational and administrational limitations in developing countries also make it often difficult to take full advantage of the reprographic techniques available for information centres. Commonly, their services are supposed to include the gathering of information, its processing and dissemination. In this context, it will have to be considered which of the existing reprographic techniques could best be applied for the benefit of developing countries.

Gathering information in a developing country most often means receiving desired particulars from abroad. Concerning the text of whole books and serials, the original publication is still the most desireable one, even where micro-publications might be available. Microcards may serve as a typical point in case. Some years ago, their distribution was rather popular for some time, paricularly from various American channels. Considering the great need for specialized information, such as atomic energy reports, medical research papers and similar publications for the United States and other industrialized countries, their libraries and information services had to be equipped with special microcard readers for that purpose. The organizational changes and capital investment involved were certainly worth it, considering the large number of microcards required within a short time. After a while, this particular technique proved to be not really satisfactory and was largely replaced by the transparent microfiche. In this way, microcard-readers rapidly became obsolete, not without having served their purpose while it lasted. Developing countries, however, had caught up with the microcard technique only at a later stage and to a much lesaer extent, hereby channelling capital investment into a doubtful economic undertaking.

In spite of that hardly avoidable experience, microcopies prove a most useful tool for gathering information from irregular sources. This is especially true for developing countries, the literary resources of which are mostly limited as far as scientific and technical information is concerned. Usually, they also happen to be situated at a geographical distance from the main centres of information. They therefore have to rely on air-mailed communications incurring considerable postal charges. In these cases, microcopies prove an ideal medium. Learning the lesson, the information centres of developing countries ought to be careful in evaluating microfiche and its latest stages: Super- and Ultrafiche for their purpose. The whole of this field is still in a rather fluid state of development. Besides, here again, a specific consideration, typical for developing countries, should be mentioned. It appears, namely, that microfiche readers can be marketed at a much more attractive price than the usual microfilm readers. The price of these microfiche readers proves reasonable indeed in relation to the high economic standard of the countries of origin. In consequence, the microfiche system becomes rapidly popular there. The same reader, however, becomes rather expensive when marketed in countries importing their equipment with heavy duties attached, in addition to freight and insurance charges involved. The considerably enlarged price must now be related to the economic standard prevailing at the reader's destination which, in the case of one of the developing countries, will compare unfavourably with that of the country of origin. Even for that reason alone, microfiche cannot be expected to become as popular in the developing countries as in the industrialized ones in the foreseeable future, notwithstanding its particular advantage for information services as such.

With due respect, therefore, to the care needed in choosing appropriate reprographic media for gathering information in developing countries, the advantage of microcopies cannot be overstressed. At the same time, it ought to be realized that these copies have in the first place always been looked upon as a medium to be scanned for the importance of its contents. Microfilms are not supposed to be read for any length of time. Instead, enlargements should be requested of frames of particular interest which, by way of scanning the film, appeared to be important. Microcopies as such remain the cheapest means for the acquisition of information and its storage,- more specifically for developing countries. Portions of particular interest, however, should be made available to the information centre's customer, where possible, by way of paper enlargements.

For processing of information and its retrieval, microcopies prove also a medium of advantage, e.g. where the classification of the information can be included in the microcopy at the time of its production. Once more, the microfiche has a good point in case. Much of the information required for retrieval can be included at the head of the fiche, where it can be read by the naked eye. Such a fiche, put in an envelope for its protection, can actually serve as a kind of index-card and source of informationat one and the same time; because, once taken out of the card-file, the fiche serves as the actual material one has been looking for. Here again, information services operating in developing countries may find themselves in a less fortunate state, except where they practically act as the producer of the microcopy involved. However, where microcopies are received from abroad without classification, alternative solutions may be available. If the copy arrives in the form of rollfilm, the lengthy material may be cut into strips, respectively of 10 single or 5 double frames. A heading is attached above each strip, which has been written by typewriter and therefore is readable with the naked eye. Formerly, for this kind of strip with heading attached, special envelopes had been readily available. Unfortunately, it appears that they have gone somewhat out of fashion; but they may still be a solution for developing countries. An improved and mechanized version of this technique can be found in Kodak's "Microstrip". This system, in turn, is largely overshadowed nowadays by the more sophisticated systems of data processing and storing. In most cases, the latter ones will prove, though useful, much too expensive for the limited possibilities of developing countries as against their extensive use in industrialized ones, for which these complicated and costly systems were originally conceived.

Generally speaking, the dissemination of information should preferably be handled by way of paper copies. These are produced from original sources available at the information centre. No real difference exists between the techniques used for that purpose in developing or industrialized countries. The great number of copier models, their materials and prices, however, make it essential for information services in developing countries to choose the appropriate product more carefully. Developing countries, usually being limited by their resources of foreign currency, even may have to give preference to techniques based on materials locally available, including advanced applications for diazo copying. In any case, the number of copies needed in the foreseeable future should be taken as basis of calculation for investment in a particular product; and although each of these models may be capable to serve for many years to come, another typical difference between industrialized and developing countries exists. In the former ones, the chosen model probably returns the investment through its use during two or three years. By that time, a more advanced model - or possibly new technique - may become available which would make it desirable to change the existing equipment, though still usable, against new one. Such a change would be economically sound. In a developing country, rentability may be less obvious after the same period. In that case, a change of model might still be desirable but without economic justification. It is for these reasons, that the choice of the appropriate copier must be made much more carefully in the case of developing countries.

Similar considerations exist where a large number of copies from the same document has to be distributed. These copies may be produced either by one of the duplicating processes, such as small-offset printing and spirit duplicating, or by making use of an additional counter at an automatic Xerox copier which makes it possible to photograph, as from the sixth copy, at a considerably reduced price. Weighing these and similar techniques may become of decisive importance for the economics of information services in developing countries; because, even economizing in a small way may, in the long run, produce important results. Limited editions of, say ten to fifty copies are typical for this kind of work. From there onwards, small-offset printing probably remains the only solution up to the stage where full printing processes become justified.

Where small-offset printing has been chosen for duplicating, a decision about preparing the offset master must yet be taken. Specialized, high quality equipment can be bought for that purpose. In developing countries, however, one may want to take advantage of electrophotographic copying machines already existing, in order to produce masters in a cheaper way with satisfactory results. Actually, the whole of these dissemination problems requires not only reprographic knowledge and experience but also an understanding of distribution matters and publishing of limited editions which are so typical for the requirements of information services. In industrialized countries, these centres of information may consider as their foremost duty to answer requests received from interested parties. In developing countries, the same centres may have to go out of their way in order to stimulate local industry, for instance, into testing new possibilities gathered from publications abroad. Suggestions of this kind may have to be sent at random to all concerned. The distribution of a few dozens of copies is therefore typical for the needs of developing countries.

Considerations like shortage of foreign currency may stimulate the setting-up of local factories for the production of certain reprographic material and, possibly, even some of the simpler models of equipment. Progress made in this direction in India, for instance, could serve as a guide. Nonetheless, careful economic calculations should show if local manufacturing would indeed prove of advantage. On the other hand, a case such as the local production of reasonably priced microfiche readers could become almost decisive for the introduction of new reprographic techniques in the service of information as pointed out earlier. Instead of designing basically new equipment, arrangements with manufacturers in industrialized countries by payment of royalties may be preferable. Even assembling parts manufactured abroad may bring about some valuable savings in foreign currency. Unfortunately, developing countries are usually also those which suffer from lack of the necessary number of technicians. This shortage makes itself felt even in the proper handling and upkeep of the more sophisticated equipment, as required by information centres. It must be realized that, although the famous slogan "Any office girl can do it" may basically be true, it is the well trained and experienced technician who must be available to the untrained personnel in order to overcome unavoidable short stoppages in the smooth running of the machines. This technical point becomes more apparent in developing countries, as the manufacturer's own technicians or those of his representative may be stationed far away. Any information centre based on reprographic services without the assurance of adequate technical help will certainly find itself limited in discharging its duties.

It would be wrong, however, to consider the utilization of reprographic methods from a technical point of view only. After all, information services are run also in a certain organizational and administrative framework. This framework in itself must be in a certain relation to the country's general organizational capabilities and its administrative network. These again may yet be less developed among the younger nations. For that reason, reprographic techniques which satisfied in the framework which they were intended for, may not be applicable for developing countries. For the solution of this kind of problem a local trainee ought to be sent abroad, preferably to a smaller one of the industrialized countries. On his return, tho by now trained professional, experienced in the local difficulties of his native country, will become the appropriate person to find the desired solution. In any case, solutions - although based on the experience gained among industrialized nations - have to be adjusted to local habits and possibilities. Prestige requests for sophisticated equipment must also be avoided as long as its true application cannot be utilized. Modern reprographic equipment, often based on electronic machinery, is typical for this kind of consideration.

With due care, however, reprographic means will prove a powerful aid to running modern information services even in developing countries. With reprography's aid, centres of information can play their proper role in the international network of information services. Through this cooperation,- not only developing countries, but the industrialized nations too, may gain important advantages. It is well known that experience gained in developing countries may be of significance even for the industrialized nations as far as similar physical conditions, for instance, exist. The language used in announcing the results of these experiments may not be known outside their country of publication. Information services would then be looked to for translating the main features of these results for the benefit of interested parties abroad and for disseminating them accordingly. In such a connection, reprography will prove a decisive medium for international cooperation.

COMPUTER TYPESETTING WITH AN EXTENDED FONT OF CHARACTERS

Horst Zuchel
Zentralstelle für
maschinelle Dokumentation(ZMD)
Frankfurt(Main)
Germany

SYNOPSIS

The limited font of characters of the used machines is one of the main problems in the field of non-numerical data processing. The example of the production of the "Deutsche Bibliographie" should shown, how the Center for Mechanisation in Documentation has solved this problem.

Im Jahre 1964 begann man bei der Deutschen Bibliothek in Frankfurt(Main) damit, die Möglichkeiten des Einsatzes der elektronischen Datenverarbeitung bei der Herstellung der Deutschen Bibliographie zu prüfen. Die ständig steigende Anzahl von Veröffentlichungen und der Personalmangel erforderten geeignete Maßnahmen, um den Erstellungszeitraum der Bibliographie, der damals z.B. für ein Halbjahresverzeichnis bei 16 bis 18 Monaten lag zu verkürzen. Die Zentralstelle für maschinelle Dokumentation(ZMD) entwickelte ein System, welches alle Arbeiten, die zwischen der Titelaufnahme und dem Druck sämtlicher Verzeichnisse und Register liegen, von einem Computer ausführen ließ. Als output lieferte dieser Computer einen TTS-Lochstreifen zur Steuerung von Bleisetzmaschinen.

So erschien am 6. Januar 1966 als erste in der Welt eine National-Bibliographie, die mit Hilfe eines Computers erstellt worden war. Damit war das grundsätzliche Problem gelöst, Bibliographien, Register, Kataloge u.a. in typographisch ansprechender Form maschinell herzustellen.

Wenn dieses erstmals praktizierte System nicht frei war von Mängeln, so waren diese Mängel in erster Linie zurückzuführen auf gewisse Unzulänglichkeiten der eingesetzten Maschinen.

Das zeigte sich schon bei den für die Datenerfassung verwendeten Lochstreifen-Schreibmaschinen, die bei 44 Tasten in Groß- und Kleinumschaltung lediglich die direkte Darstellung von maximal 88 Zeichen zulassen. Bei einem derart geringen Zeichenvorrat ist vielleicht noch die Herstellung einfachster Kataloge möglich, nicht aber die Herstellung einer National-Bibliographie. Das Problem des begrenzten Zeichenvorrats stellte sich auch bei der eigentlichen Datanverarbeitung, denn der benutzte Computer IBM 1460 arbeitet im BCD-Code. Im BCD-Code sind nur 64 Zeichen direkt darstellbar.

Den gestellten Ansprüchen nicht gerecht wurde auch die für die Ausgabe eingesetzte Bleisetzmaschine LINOQUICK. Als nachteilig wurden insbesondere die im Verhältnis zur Verarbeitungsgeschwindigkeit des Computers geringe Setzgeschwindigkeit, die unzureichenden Möglichkeiten der typographischen Gestaltung und die relative Häufigkeit von Setzfehlern empfunden. Schließlich waren auch Silbentrennung, Fahnenkorrektur und Umbruch durch den Computer noch nicht möglich.

Deshalb entwickelte die ZMD in enger Zusammenarbeit mit der Deutschen Bibliothek ein verbessertes System für den Computersatz. Die Verbesserungen reichen von einer mehr analytischen Datenerfassung über programmtechnische Verfeinerungen und Erweiterungen bis hin zum Einsatz einer ungleich leistungsfähigeren Photosetzmaschine.

Die Grundkonzeption dieses verbesserten Systems möchte ich Ihnen im folgenden kurz darlegen. Dabei lassen Sie mich Ihnen zunächst den organisatorischen Ablauf des Projektes erläutern, um anschließend zu zeigen, wie die beschriebenen Mängel und Nachteile des alten Systems beseitigt wurden.

Die Erfassung der Daten erfolgt auf 8-Kanal-Lochstreifen. Diese Daten werden off-line mit Hilfe eines Konverters zuerst auf ein Magnetband übertragen, um die Eingabegeschwindigkeit in den Computer zu erhöhen.

Bevor mit der eigentlichen Verarbeitung begonnen werden kann, muß das Material bestimmten Gültigkeits- und Plausibilitätskontrollen unterzogen werden. Die vom Prüfprogramm festgestellten Fehler werden mit einem Korrekturprogramm beseitigt. Dieser Zyklus wird solange durchlaufen, bis das Material fehlerfrei ist.

In der nächsten Phase wird das Material für die weitere Verarbeitung aufbereitet. Das bedeutet vor allem Zeichenumwandlungen, Textumsetzungen (z.B. Herausziehen von Autoren, Deskriptoren u.a. für die Erstellung von Registern), Sortieraufbereitungen und anschließendes Sortieren oder Mischen der Daten.

In der Ausgabephase schließlich werden die Daten für die Setzmaschine zusammengestellt. Das heißt: Zeilenaufbereitung mit Silbentrennung, Umsetzung des Computer-Codes in den Code der Setzmaschine und Einfügen der Steuerzeichen für den Photosatz.

Das Material wird dann abgesetzt auf Photopapier. Jede Zeile ist in dieser Phase noch mit einer laufenden Nummer versehen. Anhand dieser Nummer wird in einem weiteren Programmlauf die Fahnenkorrektur durchgeführt. Sind die Daten fehlerfrei, erfolgt der Umbruch und schließlich der Satz des umbrochenen Materials auf Film. Damit endet die Verarbeitungskette durch Computer bzw. Setzmaschine.

Folgende Maschinen und Geräte werden verwendet:
Für die Datenerfassung wird eine Lochstreifenschreibmaschine vom Typ VONAMATIC benutzt; die Konvertierung vom Lochstreifen auf Magnetband erfolgt auf einem Konverter GIER 3000. Der eingesetzte Computer ist eine IBM 1460 (16K) mit angeschlossenen 6 Magnetbandeinheiten, 5 Magnetplatteneinheiten, Schnelldrucker und Kartenabfühl- und -stanzeinheit. Für die Silbentrennung und die Umsetzung in den Code der Setzmaschine wird eine IBM 1130 (16K) verwendet. Die Setzmaschine schließlich ist eine Photosetzmaschine vom Typ LINOTYPE 505.

Nachdem auf der Ausgabeseite eine Maschine zur Verfügung stand, die bei einem Zeichenvorrat von rund 1000 Zeichen wohl allen Anforderungen gerecht wird, galt es in erster Linie das Problem des begrenzten Zeichenvorrats bei der Datenerfassung und -verarbeitung zu lösen. Die bereits erwähnte Lochstreifenschreibmaschine hat einen Zeichenvorrat von 88 Zeichen. Sicherlich gibt es Erfassungsgeräte, mit denen mehr Zeichen darstellbar sind. Es gibt aber bisher kein Gerät, mit dem alle gewünschten Zeichen zu erfassen wären. Auf der anderen Seite ist ein erhöhter Zeichenvorrat bei der Erfassung nur dann sinnvoll, wenn diese Zeichen auch zu verarbeiten sind. Das ist direkt bei dem Zeichenvorrat von 64 Zeichen der IBM 1460 nicht möglich. Auch der Einsatz etwa einer Byte-Maschine würde dieses Problem nicht lösen, denn hier liegt die Begrenzung bei 256 Zeichen. Dieser Zeichenvorrat würde zwar für die Herstellung der Deutschen Bibliographie ausreichen. Da aber das Problem des begrenzten Zeichenvorrats ein generelles Problem in der nicht-numerischen Datenverarbeitung ist, d.h. die Maschinen lassen die direkte Verarbeitung aller gewünschten Zeichen nicht zu, mußten andere Wege beschritten werden.

Eine generelle Lösung dieses Problems hat die ZMD mit der Entwicklung der sogenannten Protypen gefunden. Protypen sind provisorische Typen, die nicht vorhandene Zeichen ersetzen und erst bei der Ausgabe durch die Setzmaschine auf die ursprünglichen Zeichen zurückgeführt werden. Dadurch ist es möglich, alle gewünschten Zeichen ohne Informationsverlust zu erfassen und zu verarbeiten, auch wenn die verwendeten Maschinen nur einen begrenzten Zeichenvorrat haben. Das Prinzip, bestimmte nicht vorhandene Zeichen durch vorhandene Zeichen zu ersetzen, ist im Druckereiwesen schon lange bekannt. Die ZMD hat dieses Prinzip erweitert und in eine systematische Ordnung gebracht. Das hat die Verarbeitung dieser Zeichen im Computer erleichtert bzw. überhaupt erst ermöglicht.
Doch wie sehen diese Protypen aus? Wie sind sie strukturiert?

Eine Protype besteht grundsätzlich aus drei Zeichen: einem Grundzeichen (Grundbuchstabe oder Grundziffer) und zwei Folgeziffern. Zur Kennzeichnung dieser drei Zeichen als Protype wird ein Steuerzeichen (z.B. §) vorangestellt. Ein nachfolgendes Steuerzeichen ist nicht erforderlich, da eine Protype immer dreistellig ist.

Wir kennen zwei Arten von Protypen:

1. Protypen zur Darstellung von Sonderbuchstaben.
   Die Protype besteht aus einem Grundbuchstaben und zwei Folgeziffern;
   z.B. Sonderbuchstabe å wird dargestellt als

§a22
→ Steuerzeichen
→ Grundbuchstabe
→ Folgeziffern

2. Protypen zur Darstellung von Sonderzeichen.
   Die Protype besteht aus einer Grundziffer und zwei Folgeziffern;
   z.B. Sonderzeichen % wird dargestellt als

   §364
   § → Steuerzeichen
   3 → Grundziffer
   64 → Folgeziffern

Eine vollständige Liste aller Protypen finden Sie in der Veröffentlichung der ZMD: "Typen und Protypen für den Fotosatz mit der LINOTRON 505". Hrsg.: ZMD (ZMD-A-22). Berlin, Köln, Frankfurt(Main): Beuth-Vertrieb GmbH, 1970, 30 S.

Die Vorteile dieser Form der Zeichenverschlüsselung sind:

1. Zur Darstellung der Protypen werden lediglich die 26 Buchstaben des Alphabets und die 10 Ziffern benötigt. Das bedeutet: Protypen sind mit jeder normalen Lochstreifenschreibmaschine zu erfassen.
2. Das System der Protypen ist offen, d.h. gewünschte Ergänzungen durch den Benutzer sind möglich.
3. Die Struktur der Protypen ist immer gleich. Das erleichtert das Ablochen bei der Erfassung und macht, insbesondere durch die Verwendung des Grundbuchstabens, die Protypen lesbar.
4. Protypen sind sortierfähig.
5. Alle Protypen sind ausnahmslos auch als Indizes darstellbar. Zu ihrer Kennzeichnung ist lediglich ein zusätzliches Steuerzeichen notwendig (z.B. Über- oder Unterstreichungszeichen).

Diesen Vorteilen steht als Nachteil eine quantitative Erweiterung des zu erfassenden Materials gegenüber, denn es sind zur Darstellung eines Zeichens vier Zeichen abzulochen. Doch wird man diesen Nachteil sicherlich in Kauf nehmen, wenn es darum geht, alle Daten ohne Informationsverlust zu erfassen und zu verarbeiten.

Nachdem das Problem der Vollständigkeit bei der Erfassung der Daten durch die Verwendung von Protypen gelöst war, galt es, das Material so zu erfassen, daß ein weitgehend direkter Zugriff zu den einzelnen Daten möglich ist. Dazu ist es notwendig, diese Einzeldaten zu kennzeichnen. Eine Kennzeichnung der Daten ist grundsätzlich möglich durch Steuerzeichen oder durch die Verwendung von Kategorien. Die Deutsche Bibliothek hat ein Erfassungsschema erarbeitet, das die Benutzung von 99 unterschiedlichen Kategorien zuläßt. In der ersten Phase der Verarbeitung durch den Computer werden zunächst die Adressen dieser Kategorien (Adresse = Stelle innerhalb der Titelaufnahme) in einem Adreßbuch zusammengestellt. Bei einem derart analytischen Erfassungsschema ist dann über das Adreßbuch ein direkter Zugriff zu den Daten möglich.

Die programmtechnischen Verbesserungen des neuen Systems beziehen sich in erster Linie auf die Ausgabephase, in der die Daten für den eigentlichen Computersatz auf der LINOTRON 505 zusammengestellt werden. Es war das Ziel, die beim Bleisatz notwendigen manuellen Eingriffe weitestgehend auszuschalten und die Möglichkeiten der Photosetzmaschine voll auszunutzen.

Als Vorteile dieser neuen Konzeption seien genannt:

1. Höhere Setzgeschwindigkeit
   Eine Aussage über die tatsächliche Geschwindigkeit einer Setzmaschine fällt schwer, denn die Setzgeschwindigkeit ist u.a. abhängig von der Anzahl der belegten Magazine(Bleisatz) bzw. Rahmen(Photosatz), der Schriftgröße und der Zeilenlänge. Geht man davon aus, daß nur ein Magazin bzw. Rahmen belegt ist, daß die Schriftgröße 8 Punkt beträgt und die Zeilenlänge 50 Zeichen nicht überschreitet, so setzt nach unseren Erfahrungen die Bleisetzmaschine LINOQUICK ca. 3 bis 5 Zeichen und die Photosetzmaschine LINOTRON 505 ca. 70 bis 80 Zeichen pro Sekunde. Beim Photosatz wird also etwa die 20-fache Setzgeschwindigkeit erreicht.
   Dadurch ist die Diskrepanz in der Geschwindigkeit zwischen dem Computer und der Setzmaschine zwar nicht beseitigt, aber doch verringert. Außerdem wird der Computer zeitlich weniger belastet, denn statt des bisherigen TTS-Lochstreifens wird jetzt ein Magnetband zur Steuerung der Setzmaschine erstellt.

2. Bessere typographische Gestaltung
   Auf der Bleisetzmaschine LINOQUICK wurde nur in einer Schriftgröße gesetzt. Das führte zu einer gewissen Eintönigkeit des Schriftbildes. Die LINOTRON 505 kennt 15 unterschiedliche Schriftgrößen. Per Programm werden die gewünschten Schriftgrößen für Überschriften, Text, Durchschüsse,

Fußnoten, Nebeneintragungen u.a. vergeben und der LINOTRON 505 über Steuerbefehle mitgeteilt. Auch wurde eine Verbesserung des Schriftbildes durch eine maschinell durchgeführte Silbentrennung erreicht.

3. Ausschaltung manueller Eingriffe
   Manuelle Eingriffe an der Bleisetzmaschine sind immer dann notwendig, wenn ein Zeichen gesetzt werden soll, das nicht erfasst werden konnte. Entweder die Maschine läuft auf einen Stop und der Setzer muß das gewünschte Zeichen in die Maschine eintasten oder das fehlende Zeichen wird nachträglich hineinkorrigiert. Da nur in einer Schriftgrösse abgesetzt wurde, mußten auch die Überschriften manuell gesetzt werden. Schließlich waren auch die notwendigen Fahnenkorrekturen und der Umbruch vom Setzer durchzuführen.
   Alle diese manuellen Eingriffe sind jetzt nicht mehr notwendig, denn es sind bei Verwendung der Protypen alle Zeichen zu erfassen und deshalb per Programm bei der Ausgabe direkt in den einzelnen Rahmen der LINOTRON 505 anzusteuern. Die Überschriften werden vom Programm in den Text eingefügt. Auch die Korrektur an den abgesetzten Fahnen und der zweispaltige Umbruch einschließlich des Setzens einer lebenden Kolumne und des Auftreibens der Spalten werden vom Computer ausgeführt. Das verringert nicht nur die reine Absetzzeit ganz erheblich, sondern hilft auch, Fehler zu vermeiden, die vor allem beim manuellen Einfügen der Überschriften immer wieder vorkamen.

Zusammenfassend läßt sich sagen: Die Verwendung der Protypen und die programmtechnische Berücksichtigung der maschinellen Gegebenheiten der Photosetzmaschine erforderten einen nicht unerheblichen zusätzlichen Programmieraufwand. Doch wird die Frage des Aufwandes bedeutungslos, wenn man an die beschriebenen Vorteile denkt. Ich darf daran erinnern, daß es bei dieser Form des Computersatzes ein Zeichenvorratsproblem nicht mehr gibt, daß sich die Vorteile in der Setzgeschwindigkeit und den umfangreicheren typographischen Gestaltungsmöglichkeiten zeigen und daß schließlich manuelle Eingriffe nicht mehr notwendig sind, da Silbentrennung, Fahnenkorrektur und Umbruch per Programm durchgeführt werden.

COMPUTER TYPESETTING WITH AN EXTENDED FONT OF CHARACTERS

Horst Zuchel

Translated by the author from the German original

In 1964, the German Library in Frankfurt started to check the possibilities of using an electronic data processing machine for production of the German Bibliography. The constantly increasing number of publications and the staff needs required appropriate steps in order to shorten the period of production of the Bibliography. At that time, the period of production was about 16 until 18 months. The system developed by the ZMD, the Center for Mechanization and Documentation, allowed to process data by computer from the data recording until the printing phase. As output, the computer delivered a six-channel TTS-punched paper tape for steering a line casting machine. So, the German National Bibliography was the first national bibliography in the world produced with the aid of a computer. It was published on January 6, 1966.

Therewith, the fundamental problem was solved in a typographically appropriate manner to produce automatically bibliographies, registers, catalogs, etc. But this system wasn't quite satisfactory. First of all, it used machines that had to be regarded as weak points. That would be found at the punched paper tape typewriter used for data recording. Usually these machines have 44 keys in upper and lower case which allows for the representation of 88 characters only. Such a limited font of characters seems to make it impossible to produce a national bibliography which could meet the approval of all users. The problem of the limited font of characters was also found at the stage of real data processing, because the computer used (an IBM 1460) has the BCD code. Only 64 characters can be represented in this code. The linecasting machine, LINOQUICK, used in the output stage, also did not satisfy the required demands.

The speed of typesetting is too low in relation to the speed of the computer. Also the possibilities of hyphenation, page make-up, galley-proof correction and of an appropriate typographical layout are limited.

Therefore, the ZMD had developed, in cooperation with the Deutsche Bibliothek, a more efficient system of computer typesetting.

The improvements range from a more analytical data recording and an extension and refinement of the programming system up to the employment of a photo composing machine. I want to present this more efficient system to you in the following.

First of all, I want to point out the organizational running off and then I want to show you how we have eliminated the above described weak points.

The data are recorded on 8-channel punched paper tape. These data are converted on a magnetic tape, in order to increase the input speed for the computer. Before we can start the real processing, we have to check and to correct the data. Then it is necessary to prepare the data for the following program. That means character transformation, permutations (for example, extract of authors, descriptors for producing registers), preparing for sort and the following sort or merge of data.

In the output stage the data are prepared for the typesetting machine. That means line preparing and hyphenation, changing the computer code into the code of the typesetting machine, and to insert the control signs for the photo composition machine. Then the data are set up in type on the galley proof. Each line has a current number for proof corrections. If the data are flawless, we can do the page make-up and finally, the typesetting on film. This is the end of the processing chain.

The photo composition machine, LINOTRON 505, has a font of about 1000 characters. That satisfies all required demands.

Now, it was necessary to solve the problem of the limited font of characters at the data recording and processing. The mentioned paper tape typewriter has a font of 88 characters, I am sure that there are recording machines which can represent more than 88 characters, but I am also sure that there is no machine which can represent all the characters you want. On the other hand, an extended font of characters at the data recording is only effective if you can process these characters. That is impossible with a font of 64 characters of the IBM 1460. Also the use of a type-machine wouldn't solve this problem, because the limitation of this machine is about 256 characters. Of course, this font of characters would be sufficient for production of the German Bibliography. But as the problem of the limited font of characters is a general problem in the field of non-numerical data processing, it was necessary to find a way to extend the available font of characters as comprehensively as possible.

The ZMD established systematic tables where special characters are represented as so-called "Protypes". Protypes stand for the original characters which cannot be otherwise represented during certain processing pahse, e.g. especially during data recording. In principle, the method of replacement is well known in the printing business. But in contrast to the old fashioned replacement methods, ZMD arranged comprehensive collections of special characters and organized them systematically in tables. The representation chosen takes into consideration a simple handling for data recording and the further computer processing.

Seen in detail protypes consist of three characters: the basic letter or figure and two following figures. Protypes have to be marked in the text by a preceding control character, e.g. § . There is no necessity to mark the end of the protype because of its fixed length.

There are two kinds of protypes: 1) protypes for representation of special letters. For example, the special letter, à, is represented as §a 22, with the paragraph sign as the control character for the protype, a as the basic letter, and 22 as the following figures. 2) Protypes for representation of special characters. e.g., the special character %, is represented as § 364 with the paragraph sign as the control character, 3 as the basic figure, and 64 as the following figures.

The advantages of this kind of character encoding are:

First: you need for representation of protypes only the 26 letters and the 10 figures. That means protypes can be types, without exception, with any normal punched tape typewriter.

Second, the system is open ended and further diacritics may be added and any corresponding letter with its two-digit number has already its fixed place.

Third, protypes are always built in the same way no matter what they are used for.

Fourth, all protypes can be used without exception as indices. Only one additional control character must be inserted between the first control character and the protype. The inserted control character indicates whether the following protypes is in superscript or in subscript position. The additional control character is represented as overlining or underlining with a not movable type-bar of the typewriter. The protype itself remains unchanged. It is important for a simple handling of recording and for the proof reading.

The disadvantages of this system have to be seen in a quantitative extension of the recorded data, because it is necessary to type four characters for the representation of one character, but certainly, you would put up with this disadvantage if you have to record all data without loss of information.

The improvements of the program were, first of all, the output stage. It was the aim to eliminate the manual handling by using a linecasting machine and to make the best of the possibilities of the photo composition machine.

The advantages of this system in using a photo composition machine are:

1. Higher speed of setting. It is very difficult to give a generally valid judgement on the real speed of a typesetting machine. The speed depends on the number of covered magazines (by linecasting) or frames (by photo composition), the size of font and the length of the line. For example, if only one magazine or frame is covered, the size of the font may be 8 points and the length of a line doesn't exceed 50 characters, we have found that the line casting machine, LINOQUICK, sets up in type about 3 to 5 characters and the photo composition machine about 70 to 80 characters per second. So we can say the photo composition machine is 20 times faster than the linecasting machine.

2. More extensive typographical layout: the line casting machine LINOQUICK sets up in type in one size of font only which brings about a rather monotonous layout. The LINOTRON 505 has 15 different sizes of font. So we can in the program have the heading, the text, the spaceline, the footnote, etc. set up in the type all in different sizes of font. Furthermore, we have improved the typographical layout by automatic hyphenation.

3. Elimination of manual handling: manual handling on the linecasting machine is necessary, if there is a character to set up in type which could not be recorded. It is also necessary for setting up headings (titles), for proof-correction and for

pagination. All this manual handling is superfluous now in the new system.

Finally, let me say: All desired characters we can record and directly set up in type on the photo composition machine. The headings are inserted into the text as well as the galley-proof correction and the page make-up are put through by program. That reduces not only the setting up time but also it means the avoidance of typical errors made in hot-metal typesetting.

MECHANIZATION OF PUBLICATIONS

M.D. Martin, P. Simmons, M.J. Vernon
(INSPEC, The Institution of Electrical Engineers, London)

SYNOPSIS

INSPEC's integrated computer-based production system for its publications is described in three parts: the background to the developments undertaken during the past four years; an overview of the system as it now operates; and a survey of some of the problem areas which have been encountered in day-to-day running.

On January 2nd 1969 The Institution of Electrical Engineers (IEE) published the first issue of Physics Abstracts to have been produced by computer-controlled phototypesetting techniques. Since that date, the whole range of INSPEC publications - abstracts journals, indexes, and titles bulletins - has been compiled and produced as part of an integrated computer-based system.

The present paper describes INSPEC's experience with this system, concentrating particularly on those aspects which concern the production of printed publications, though it must be remembered that this is only one of a number of facets of the total system.

The development of the INSPEC production system

Since 1898, the IEE has published Science Abstracts, which today comprises three abstracts journals covering physics, electrical and electronic engineering, and computers and control. Each section has an associated titles publication under the generic name Current Papers. The abstracts are of papers selected from the periodical literature and from conference proceedings, books, reports, patents and dissertations. The present volume of publication amounts to some 140,000 abstracts per year.

The abstracts are published within a subject classification having some 600 chapters and sub-chapters, and each is indexed by subject under some 3000 controlled headings, with free-language entries, to enable a six-monthly index to be prepared. Each issue of the abstracts journal carries an author index, which is also cumulated each six months; and a number of other indexes are provided. The titles publications are classified to the same level as the abstracts journals, but have neither abstracts nor indexes, and are designed to serve as a current-awareness tool for individual scientists and engineers.

The decision to mechanize the production of all these publications was taken early in 1967, following a series of research studies. This decision would probably not have been taken - certainly not at such an early date - if it had been viewed solely as the mechanization of a conventional publishing operation, for it cannot be claimed to have reduced the printing and publishing costs, taken in isolation. Nor, on the other hand, has it significantly increased them; and it must be borne in mind that the break-even point between conventional print production and computer-controlled phototypesetting, in the context in which INSPEC operates, is almost certainly getting closer year by year.

The importance of the decision to mechanize the INSPEC publications was that it represented a commitment to develop a total computer-based information service, of which printed journals would be only one of a number of products. Other products, such as magnetic tape services and SDI, are outside the scope of this paper, but they have been considered as an integrated part of the system from the very beginning. Their implementation has been phased over the past two years, and is continuing at the present time.

The heart of the INSPEC production system is a single data base into which are entered bibliographic records covering the whole subject field of the services which INSPEC provides. Each record contains an abstract, bibliographic description, classification and subject indexing. From this single data base the required subsets are selected and processed to provide a variety of different outputs.

System overview

In outlining the production system as it now operates, it is convenient to describe the procedures used for journal literature, which forms over 70% of the data base. For other types of literature, the system is similar in principle but differs in detail.

At acquisition, the receipt of the journal issue is recorded, and a control sheet is attached to the contents page. The issue is marked out to the particular group of information scientists which regularly handles this journal, and an identification number is entered on both the control sheet and the contents page. The issue then passes to the group leader responsible, who marks up his selection on the contents page, by entering one or more of the letters, A, B or C against each title, to indicate whether the item is required for Physics Abstracts, Electrical & Electronic Abstracts or Computer & Control Abstracts.

Almost all English language material, or material carrying English abstracts, is dealt with in-house. Some foreign-language journals are handled completely by a single outside abstractor; others are dealt with by sending individual papers to abstractors.

For English-language material, the information scientists classify and index the paper (entering the classification numbers and index entries on a work sheet). They may compile an original abstract or edit, abridge or augment the author abstract, as required.

After the editing stage is complete, the issue is returned to the acquisitions section to be marked off against their records, and the copy prepared for keyboarding. Wherever possible, the original title page of the paper, torn from the journal, is used as copy for the keyboard operators, who themselves pick out the various data elements such as title, authors' names, abstract etc.; so that transcription is reduced to a minimum. The information scientists' worksheets are attached to the corresponding title pages, and the batch of copy is headed by the journal control sheet. The contents page is filed as a historical record.

At the present time, all input to the data base is keyboarded on to paper tape, using specially modified Flexowriters. No capital letters are used; instead, the ribbon colour shift is used and capitals appear as lower-case letters in red. The upper-case positions on the type bars are used for common Greek and mathematical symbols, while other special symbols are entered by means of an escape code.

The paper tapes are run on to the computer and added to a "Current File", of material awaiting publication. At the same time, validation checks are applied by computer program, and a proof listing with error reports is produced. Since the line printer has a character set of only 64, it is necessary to represent other characters in a special way. In general, this is done by putting the name of the character (or an abbreviation) between a pair of delimiters. Thus Greek γ is printed as #GAMMA# .

The computer-produced proof is read by a specially-trained staff, and corrections are marked on the printout. Papers which require no corrections are immediately released for production use; i.e. there is no need to wait for a complete primary journal issue to have been corrected before releasing individual papers.

Corrections are keyboarded on to paper tape, and merged with the Current File as part of a single daily update, at the same time as new items are added.

Records are released into the data base by directives on paper tape which are also entered during the same update run. The records remain physically on the Current File until a journal production run, at which point they are removed from the file and processed for publication.

The production of an issue of one of INSPEC's Science Abstracts journals starts at the point when the material for that issue is already in existence on the Current File, having been input, corrected and released since the last issue of that journal. The processes described apply equally to any of the three abstract journals; the production procedure for the titles bulletins is essentially similar, though somewhat simpler.

The first computer program in the production suite selects material from the Current File which is required for the particular issue, rejecting incomplete or uncorrected items. The Current File is structured in batches, each of which represents a complete journal issue or other source publication, so that such information as "journal title" and "date of publication" need be stored only once at the head of each batch. The next stage of processing involves sorting the journal file into classification order. At this point the file loses its former batched structure, so that it is necessary during the selection run to ensure that the information which previously appeared only in the batch header is now associated with each selected record individually.

The selected file is sorted into the classification sequence indicated by a chapter code which is assigned to each entry by the editors. Additional chapter codes are assigned for creating cross-reference entries. Within the sequence indicated by classification codes, items from the same source publication are grouped together in page number order.

The next program in the suite performs the functions of numbering each item on the sorted file and of creating a file of cross-references and an index file. The numbered journal file contains all the main entries for the abstract journal in the correct classification sequence. Each entry on this file has all the information which was included in the original Current File record, and it is this file which is merged with those from previous numbered issues to produce the archive file for retrospective searching. Each entry in the file of cross-references includes a chapter code, the title of the document, and the abstract number to which it refers. The index file includes author index entries and entries for special conference, report, patent, book and bibliography indexes which appear in each issue of the abstracts journal. Since subject indexes are produced only on a six-monthly basis, their production is handled separately. The file of cross-references is sorted into classification sequence in the same way as the main journal file. It is then merged with the numbered journal file and with a permanent file of classification headings, to create the complete text file ready for composition.

Up to this point the text files contain few of the codes needed to format the records for printing. Apart from upper and lower case shifts and occasional words or symbols in the text of the abstract which are required in italic or bold, the entries contain no style or formatting codes which, were they present, would make searching or other uses of the data more difficult or less efficient.

The formatting of the records for printing is achieved by computer program. It is done in two stages, mainly because of restrictions in the size of the computer store. The first stage combines all the data elements needed in each printed entry in the required sequence according to sets of parameters in the composition program.

The typesetting software is highly generalized, and is driven by two levels of parameters. The first level identifies the boundary conditions required between records of different types; i.e. the spacing and style changes which occur, for example, when a main heading is followed immediately by a sub-heading, or when an abstract is followed by a sub-heading. The second level identifies the data fields which are to be selected from each type of record, the sequence in which they are to be printed, and the style, spacing and punctuation which is required for each field. Fixed literal information can also be entered into the parameter list, and the parameter format allows certain types of conditional statement to be used to vary the output in accordance with the presence or absence of selected fields.

The structured records contained within the input file are thus converted into a continuous character stream including typographic function codes, and the program justifies the text, hyphenating automatically where necessary, and writes each line forward to magnetic tape. This tape is input to the final program in the suite which breaks the text up into columns and adjusts the inter-entry spacing so that each column has the same length. The column make-up program also introduces running headings and page numbers, and performs the function of translating the file into the code structure for driving the photo-composing machine. At present a Photon 713 Textmaster is being used to filmset the INSPEC publication.

As far as possible, INSPEC production software has been written in a general-purpose way in order to minimise the programming effort required to process INSPEC files, and so that the software can easily be adapted for file maintenance and typesetting work for other applications.

Problems of a mechanized publications system

There can be no doubt that all the major problems which have been encountered in over two-and-a-half years' running have fallen within what may broadly be described as the "input" end of the system, i.e. between the initial keyboarding of new items and their release for publication.

Once a clean record has been released into the data base, its subsequent processing is straightforward and, with proven programs, has turned out to be almost entirely error-free, as might have been anticipated. The only remaining source of difficulty has been in the compilation of periodic subject indexes, for two reasons: one, that it has not as yet been possible to implement a system for the complete validation of subject index headings at the input stage; and second, that indexers do still find it very desirable to be able to review their choice of terms in the light of the appearance of the final sorted index; so that it has been necessary to permit corrections to be entered even at this late stage. This is the only situation in which the system departs from the guiding principle that a clean record should result from the initial input procedure, and be processed thereafter without outside intervention.

The final output from the system to the photo composing machine has similarly been quite trouble-free. Some minor problems of tape compatibility and certain anomalies in the control codes for the filmsetter were successfully overcome at the beginning of the project. Some slight misalignment occurs from time to time as a result of electro-mechanical faults, but this is normally identified by the printer and corrected by a re-run of the relevant part of the job. It is worth pointing out that, while it is likely that the system will be transferred to a faster filmsetter in the near future, the comparatively slow speed of the Lumitype 713 (some 70,000 characters per hour) has never seemed to be a matter of real concern. The move to a newer machine will be justified on grounds of cost and flexibility, not on grounds of speed. The improvement in currency to be gained by reducing the filmsetting time of an abstracts journal issue from 1-2 days to 1-2 hours is not really significant within the overall timescale of production.

Returning to the input area in detail, the major factors are probably the difficulty of dealing with a very large character set within the limitations of conventional computer input/output devices, and the problems of automatically validating much of the type of data which the system handles.

The total INSPEC character set amounts to some 700 characters. If the distinctions between roman, bold, italic, normal, subscript and superscript are eliminated, the set reduces to just under 200 discrete symbols. It has proved possible, with quite surprising success, to handle this character set with modified Flexowriters and a wholly standard 64-character line-printer as the proofing medium; but nobody who has been associated with the work would regard this as a satisfactory long-term solution.

Current developments at INSPEC are leading towards the use of on-line visual displays for correction and release, perhaps associated with an extended character set line-printer for proofing; with some parts of the original input on-line, and others entered by off-line keyboarding on to magnetic tape. At the present time we are still sceptical of the economics and the effectiveness (particularly in terms of operator acceptability and efficiency) of using visual displays for proofreading.

It is easy to talk of using on-line displays; until recently there has been a real problem in trying to go beyond theoretical studies. In general, computer and peripheral equipment manufacturers are oriented towards a commercial data processing market, and towards large-scale batch production. There have been many display terminals on the UK market: few, if any, have been suitable for handling large volumes of scientific text. This position is changing slowly; 96-character displays are replacing 64-character displays as the norm, and hardware-generated set of 192 characters now seems to be feasible at a reasonable price.

The problem of validating bibliographic records and reducing error rates to an acceptable level is a difficult one. The fundamental conclusion to be drawn from INSPEC's experience that if an item can be validated in any way by computer program, it is worth doing. Most numeric or alphanumeric codes used in the production system carry check characters. Classification codes are specifically checked by table look-up. Some fields, such as the text of an abstract, are regarded as wholly intractable. Others, such as author names, which appear on the face of it to be almost impossible to validate, can in fact be subjected to certain simple checks concerned mainly with format and presentation. Care has to be taken in many instances to ensure that the punctuation in the input record is consistent with the punctuation which is applied by computer program at the typesetting stage. Additionally, where much of the data consists of free text, it is necessary to take careful account of things which a keyboard operator may do which would not appear to be incorrect at keyboarding time, but which would cause anomalies in the output: eg. the introduction of superfluous 'space' characters at the end of a line.

There is much scope for the use of authority lists in an input system of this kind; these are used in the present system for journal titles, classification codes and chapter headings; and at the index production stage for subject index headings. In the near future it is hoped to implement a direct access system for using the autority file of subject index headings at input; and to study the feasibility of a similar approach to dealing with frequently-cited organisation names and addresses.

Despite these problems, the present system has proved itself in operation over the past two-and-a-half years, and has been one of the first to demonstrate the feasibility of integrating the total production of a large-scale abstracting service into a single computer-based operation. After some teething problems during the first year, the system has settled down satisfactorily and currently achieves a throughput time which is notably better than the previous conventional operation, both as regards the abstracts journal issues (now averaging less than three months from the receipt of a primary document to the despatch of the printed abstracts journal) and the periodic indexes, which are normally with the printer within four weeks of the production of the final issue for the period.

COMPUTER AIDED TYPESETTING FOR THE ENCYCLOPEDIA JUDAICA

Meir Doron
Keter Publishing House, Jerusalem

SYNOPSIS

The 16-volume Encyclopedia Judaica prepared by the Keter Publishing House is being produced by computer-aided typesetting. The system, described in this paper, saves time, ensures easy updating, and avoids the need for skilled typesetters. The Index volume, which will include 200,000 entries, will be ready two weeks after the final volume is completed.

The Keter Publishing House (a subsidiary of the Israel Program for Scientific Translations) has its own typesetting, printing and binding facilities. In 1968 these typesetting facilities consisted of Monophoto (1st generation film typesetting) equipment. When the firm decided to publish the Encyclopedia Judaica (16 volumes of 800 pages each) it was realized that additional equipment would be required, especially when considering the very tight schedule which called for publication before the end of 1971. The management of the firm decided to go into computer-aided typesetting for five major reasons.

1. Possibility of easily updating future editions of the Encyclopedia Judaica.
2. The ease with which material can be sorted by computer would permit the setting of articles in random order as they come out of the editorial department without waiting for a complete volume. This would save months of time.
3. Making corrections by updating computer tapes rather than stripping in corrections by hand would solve the acute problem of lack of skilled manpower for hand corrections.
4. The possibility of automating or partially automating the preparation of the index volume would save months of time.
5. Since there were no experienced typesetters available, the problem of training a large staff in a short time would be less complicated for a computer-aided system.

The following hardware was used

1. Input was in the form of paper tape punching equipment with hard copy. Originally five Facit 6200 keyboards were purchased but these proved to be too unreliable. They were replaced by ten Friden Flexowriters.

2. The IBM 360-50 of the Mekoroth company was used to process the material on a time rental basis.

3. A Photon 713-10 Textmaster driven by a Hewlett Packard 2020 magnetic tape reader as well as its own paper tape reader, was purchased by the firm.

4. For making final corrections directly by hand, Friden LCC-VF keyboards were used.

The software was prepared by "Natam", who planned the whole system in close cooperation with Mr. Moshe Shalvi and the staff of Keter. The tight scheduling required some hectic work by the programmers to meet the deadlines.

Input

The prime requirement was to make the system as uncomplicated as possible for the inexperienced typists who had no knowledge of typesetting.

Since the hard copy was to be proof-read, it was typed on continuous stationery, each page having room for a header and line numbers from 1 to 30. The basic unit in the Encyclopedia was the article, which was subdivided into pages and lines. Control codes were identified by a square on the hard copy which was a "dead" key so that letters could be typed inside it. (On the perforated tape it appeared as a unique code) so that the control codes for italic look like this $\boxed{i}$, bold $\boxed{b}$, roman $\boxed{r}$, small caps $\boxed{c}$, 8 pt, 9 pt, etc. $\boxed{8}$ $\boxed{9}$, end of paragraph $\boxed{e}$ and many more. If the typist felt that she had made a mistake she could cancel one or several letters by hitting the back space key (a unique code), slashing out the unwanted letters, and then typing the letters correctly. The proof-reader then simply ignored the slashed letters. The typist was not allowed to break words nor was she required to keep to any line length. Accented letters were set in two codes but on the hard copy the letter appeared in the regular way. For example, "â" was set thus: "^" (a dead key) and "a". To the proof reader it appeared then actually as "â".

Ligatures were ignored by the typist but were converted by the computer.

The resulting paper tapes were read into the IBM 360-50 of Mekoroth and processed through the input stage which formatted the job, checked for wrong codes (which were drawn to our attention by a print-out of error messages), and held on magnetic tape for future updating. Meanwhile the proof-readers read and marked the hard copy.

Corrections to the hard copy

These were executed by the same typists. They accessed the proper place for the corrections by typing (along with the proper control codes) the article number, and page and line number on the hard copy. They had the option of correcting the whole line or correcting "unique phrases" which contained the word before and after the word to be corrected. The correction tapes were then sent to the computer where they went through the input program and were then used to update the original tape in the update program. The updated tape was then put through the composition program which cut the text into lines and created new line numbers. At this stage, because of the many editorial changes that were expected, the lines were still unjustified and no word breaks were made. The next stage was the conversion into Photon codes by the computer.

The resulting magnetic type was converted into type by the Photon 713-10 at a speed of about 25 characters a second. A special type matrix drum was designed for the Encyclopedia Judaica project. Since there was room for 768 different characters, we were able to include all the required type fonts, accents, and special signs. We used Times New Roman, with italic, bold and small caps, Univers Bold and Medium, Greek, superior and inferior numbers and many accented letters. Almost all accents were in the form of floating accents, so that a very few places on the matrix drum could produce a much larger number of accented letters. Our list of matrices and their different combinations actually come to more than 18000 characters.

The resulting bromide print was copied onto sun-print paper and passed onto the proofreading department and from there to the editorial department for an editorial check and possible last-minute updating.

Corrections to the unjustified first proof

They were prepared exactly as were the hard-copy proof corrections except that this time the typist used the new line numbers as generated by the computer. In addition, it was necessary at this stage to arrange the articles in the proper alphabetical order. Since the rules for alphabetizing are rather complicated it was decided, rather than writing a program, to punch IBM cards with the names of the articles and arrange them in proper order by hand. The computer then sorted out the articles according to the order of the IBM cards.

Further, the lines had to be justified now (the Photon actually did the justifying, but they had to be cut into proper length by the computer), and the IBM hyphenation sub-program hyphenated the words that had to be broken at the end of lines. This program uses a logic system as well as an exception dictionary. This program worked very well for English, but did not always break foreign words or names correctly. If the break was not correct, this particular word was then added to the exception dictionary and was therefore broken correctly from then on.

This new tape was again run off on the Photon with new line numbers down the sides (generated by the computer). The resulting justified galley proof was again passed on to the proof reading department for checking.

Corrections to the second prood (justified)

The whole procedure was repeated as above except that the output of the Photon this time was on film rather than on bromide paper.

Corrections to the film and make-up

Since it was expected that the film would be almost completely free of errors, any minor corrections still to be made would be set on justifying keyboards (Friden LCC). The resulting paper tape was put through the Photon and the correction lines stripped in by hand.

At the beginning, we planned for automatic page make-up by the computer, which will indeed be used on the index volume (see below). The inclusion of thousands of illustrations would make a fully automatic system impractical especially from an esthetic point of view. We found that manual make-up would be faster and cheaper.

Occasional changes were made in the procedure outlined above: the hard copy correction was almost completely abandoned part way through the job. It was found not always to be the same as the input tape due generally to some procedural error on the typists' part. Besides, it was found that proofreading was much more accurate when reading the actual type rather than interpreting control codes. Occasionally, due to heavy editorial changes, we introduced another round of corrections for certain articles. At other times we skipped certain rounds especially for articles arriving late.

At the writing of this paper all the material for the encyclopedia has been input, and the first ten volumes (out of 16) have actually been printed. The rest of the encyclopedia is scheduled to be completely bound before the end of the year.

The index volume

One complete volume of the encyclopedia is to be an index, containing over 200,000 items. An index of this size, prepared by conventional means would be finished many months after the completion of the rest of the encyclopedia. By using an automated system, the index volume will be ready only two weeks after the last volume is completed.

Dr. Posner, who heads the index project, together with the people at Natam, devised the following system:

1. The indexers read the manuscript and mark each word that is to appear in the index with codes containing serial numbers $\boxed{I}$ $\boxed{1}$, $\boxed{I}$ $\boxed{2}$, $\boxed{I}$ $\boxed{3}$, etc.), the entry-words for each article being numbered from 1.
2. At the same time the indexers prepare a file card, marking down the word or words to be indexed, the article number in which it appears and the serial number in that article.
3. The file cards are edited into main items, sub items, etc. and filed alphabetically in boxes. As new items come in, re-editing is often required. Several identical items may appear on one card each one with its article and serial numbers.
4. The input typists type each card in its proper alphabetical order, giving the item, article numbers and serial numbers. This operation was started a short time before all the articles were actually set, any additional entries being included as corrections (see below).
5. Proofs are read on hard copy and then on a Photon run of the computer input. All the items in this stage still appear in their unconverted form exactly as they were input. Corrections are punched and the material updated.
6. Since make-up of the encyclopedia itself was not automated, it is now necessary to "inform" the computer which article is contained in each column of the encyclopedia. (for example: column 532 contains article number 535403 from line 23 to line 78). This operation proceeds immediately upon make-up and will be finished in time for the next stage.
7. The computer then converts the serial and article numbers into volume and column numbers (the columns are numbered rather than the pages). This means then that the item is identified as in any index, the reader being directed to the proper page.
8. This is run off on the Photon, corrected in the usual way, and made automatically by a computer program, which arranges the material in four columns with running heads and page numbers automatically inserted.

MECHANIZATION AND AUTOMATION IN PUBLISHING

Lowell H. Hattery
Professor of Management
and Public Administration
The American University
Washington, D. C.

SYNOPSIS

To achieve maximum potential of technological innovation in publishing, a higher order of systems concept and man-machine relationship is required, including on-line, inter-active editing. Substantial investment must be made in systems research and software development. Information specialists should be leaders in conceptualization, research and action toward more effective mechanization and automation in publishing.

Mechanization and automation of the printing-publishing process has several significant points of impact.

Technological change in printing and publishing is of special interest to the printing industry for it alters methods of production, changes economics and affects organization and managerial processes. It is significant to publishers and to users because technological changes may affect the nature of the medium, cost per unit and time of delivery. Technology is of interest to the author for the form of his manuscript may be constrained, and because he too is interested in time, cost and output form.

Finally, technological change in printing and publishing is of interest to the information specialist since it affects the input, the store and the output of information for an information system.

Information Systems Slow to Automate

A somewhat inexplicable lag in adapting new technology has characterized information services. Despite the fact that electronic information processing machines have been available for 25 years, information specialists have been very slow to mature a concept of achieving optimal integration of new technology into operational information systems.

One of the reasons for this deficiency is the historical separation of the functions of information development, printing, storing, retrieving, reformatting, distributing, and revising the store. Even if one institution has control over these functions, occupational specialization has effectively maintained separation in thought and action.

Yet the sequence of functions can be viewed as a single process; they must be so viewed to utilize new technology optimally.

Another barrier to achievement of automated systems in information has been the absence of standards and standardization. The myriad of type faces, different forms of citation, various indexing and classification systems, spelling variations, etc., etc., all contribute to confusions in the system which complicate and increase the cost of computerization and automation.

Electronic Composition

One of the most intriguing innovations in the information process is electronic composition, also referred to as computerized typesetting.

Although application of the electronic computer to typesetting seems an obvious and straightforward process, it is just becoming a significant production factor. A dozen years of fumbling experimentation is now clarifying the nature and the economics of the process. Modification and refinements of hardware, software and production methods have been developed encouragingly in the last three years. And a better understanding of input requirements and of output potential gives a basis for optimism for the future.

Of greatest potential impact for electronic composition is a better design linkage with information stores.

A second important contributor to the growth outlook is the availability and utilization of smaller, lower-cost machines, somewhat comparable to the development and popularity of minicomputers. The editor of Reproductions Review said recently: "In the past decade the art of phototypesetting has evolved from the prohibitively expensive province of the giant publishers to a versatile tool available to both high and low volume user at either end of the price scale."[1]

In its 1970 survey of computer typesetting, Composition Information Services, Inc. reported that in the United States 1,400 computers are in use or on order for typesetting. CIS also reported in surveys of available equipment that the number of "low-priced typesetters able to accept unjustified type" has more than doubled in less than two years.[2]

1 Riley, W., "Phototypesetting Today: Where Is It", Reproductions Review, April 1971, p. 24.
2 CIS Newsletter, February 1, 1971

Printing industry consultant P. L. Andersson discusses the growth of computer typesetting in the following statement:

> The reasons for the popularity of phototypesetters are simple and direct. First, much more printing is being done today than ever before, and the higher speeds possible with electronically driven phototypesetters are needed to keep up with the market. Second, more and more printers are converting to offset lithographic printing, rather than letter press. The lithographic process uses a photographically prepared plate, and thus is more compativle with photographic typesetting than the letter press process which utilizes the output of type and line casting machines directly. Third, and more to the point, computers have gained a wide acceptance in the printing industry as machines which will hyphenate, justify, format, edit, and merge corrections and alterations.[3]

The most critical unsolved problem in electronic composition is the development of adequate software.

Efficient software is a most formidable task if the machine system is to perform in the patterns similar to manual composition. Decisions about hyphenation, etc. require a very complex set of rules and a large store of exceptions: page make-up is a complex task if the decisions are to be entirely machine-controlled.

Unfortunately, software research and development has not been given adequate support.

The requirements for a comprehensive software project are presented by John W. Seybold in the proceedings of a 1970 conference on computer composition.[4] Mr. Seybold noted the need for "general-purpose computer typesetting software which can be used on standard medium-scale computer equipment." He adds: "I do not believe that any present software is sufficient, or adequate, or general enough, or straightforward enough, or versatile enough."

Development of satisfactory general programs for computer composition will require a considerable investment of time and money. Standardization of many variations among composing conventions is highly desirable. International cooperation in standardization of conventions especially and in software development could be helpful to all.

---

[3] Andersson, P. L., "Phototypesetting -- A Quiet Revolution," Datamation, December 1, 1970, p. 22.

[4] Landau, R. M., editor. Proceedings of the ASIS Workshop on Computer Composition. Washington, D. C.: American Society for Information Science, 1971, pp. 87-98.

Composition Input

Efficient input for automatic composition is a principal concern of industry specialists.

High time and cost factors in printing and publishing reflect in large measure preparation of a manuscript and the introduction of that manuscript into composition of various possible forms. Only recently has hard attention been directed to designing input systems which are directly related to an automated concept.

Optical character recognition is an obvious approach and several machines are available.

The principal commercial use of OCR has been the high-volume processing of less than page-size documents such as credit card and charge slips. Page readers are more costly. Nevertheless, there is increasing use of OCR for publishing. P. L. Andersson estimates that if the type font can be controlled, the overall cost of reading OCR scannable copy into a typesetting system can range from 1/4 to 3/4 the cost of reading perforated paper type. Andersson is optimistic about future applications. He adds:

> On the basis of the success to date of these installations /OCR page scanners/ and on the continuing improvements in technology as well as rapidly decreasing costs, application for OCR in the graphic arts can today be considered to be in the same position as typesetting computers enjoyed in 1963 when there were only three such installations. I believe that we can look forward to a rapidly increasing machine population...new pattern recognition techniques, new direct optical computation techniques and a growing appreciation of the need for editing and correction capabilities will aid in continuing this growth.[5]

Andersson adds that "OCR must be regarded as the only feasible electronics system for releasing the graphic arts from its enslavement by the fumbling fingers of the keyboard operator."

Reformulation of the System

A private study produced for the printing industry by Battelle Research Institute was reported in a 1970 conference known as "Comprint 90". The report and the conference looked twenty years ahead. In a summary and commentary, a trade journal editor stated:

---

[5] Andersson, P. L. "Optical Character Recognition for Typesetting," in Lowell H. Hattery and George P. Bush, editors, Technological Change in Printing and Publishing Management. Washington, D. C.: Spartan Books, Inc., in press.

> Half an industry and half a communications system, the printing industry will be the first manufacturers to feel the consequences of a society dominated by information rather than commodities. As the Battelle study put it, "The printing industry will be the world's first to use the computer as its principal means of production." In this central role, the computer will replace the printing press.[6]

The versatility of computer production of packaged information is also noted by Guy Farrell: "Computerized information can be used to create a book, magazine, microform, facsimile newsletter, or video-cassette program with only variations in procedures. The possibilities for creating different products from the information resources available to diversified publishers are virtually limitless."

This flexibility serves too the increasing demand for output of publications in limited numbers of copies -- the "short-run" print-out.

Publishers are in fact developing a new perspective about their role in an information system for a "knowledge-oriented society," according to Paul D. Doebler, editor of Book Production Industry.[7] Doebler points out that publishers have begun to see that "publishing is not simply the business of producing books or magazines -- or for that matter, radio and television programs. It is, rather, the business of moving information from those who have it to those who need or want it, by whatever means apply." Doebler observes further that, "while information science developed from a library orientation, much of its essence is basically publishing rather than storage and retrieval." This concept is reflected in a new publishers' trade association in the United States which is incorporated as the Information Industry Association.

Electronic Editing

The facility for editing text for publication on a cathode ray tube display is another important element in the publishing system. One editing system, for example, displays 50 lines at a time. Corrections can be made through a keyboard and immediately appear on the screen. Lines are automatically justified. An approval signal dispatches the copy for typesetting.

---

6 Farrell, G., "The Larger Meaning of Comprint 90," Book Production Industry, January 1971, p. 32.

7 Doebler, P. D., "Publication and Distribution of Information" in Carlos A. Cuadra, editor, Annual Review of Information Science and Technology, Vol. 5. Chicago, IL: Encyclopaedia Britannica, Inc., 1970, p. 223.

One of the most exciting experiments in bringing the author closer to the machine and therefore to output has been carried out by Jerry A. Carlson, managing editor of the Farm Journal, a large-circulation magazine.[8] Carlson himself points out a long history of fruitless efforts to persuade writers to compose on Monotype machines and Teletypesetter keyboards. However, the editing possibility on these machines was negligible.

Mr. Carlson first developed a research project with Iowa State University. Within the Journalism Department a "scientific study of video-tube writing" was conducted. He reported that the study "established that a useful number of our editors would and could write and edit articles with on-line video terminals."

Although there are several ongoing examples of on-line, video screen composing and editing, the process is still in early stages of development. Carlson analyzes its usefulness for three writing situations:

1. The "quick-and-rough" writer under pressure will key his story into a video terminal, a rewrite editor will call it up onto his terminal, correct it with light pen and keyboard and send it to the computer. The article "never saw manuscript paper, paper tape, red pencils, green pencils, or blue pencils."[9]

2. The creative author, such as the book author, will key in his text. "He can just close his eyes and type, never worrying about carriage returns, jammed keys or typos. He simply empties his heart into an electronic heart miles away... Next morning, the author can scroll up through his previously written thoughts, correct the typos, repair his transitions and perform other insertions and deletions."

3. The author who must develop highly polished or concise copy "will insert, delete and revise a great deal as he goes along," according to Carlson. One of Carlson's editors commented: "The words become almost physical, plastic -- you can hammer and pound and sculpture them into shape."

Carlson summarizes his view of video editing: "... the new technology offers writers and artists wide opportunity for greater aesthetic and intellectual achievement."

---

[8] Carlson, J. A., "Video Editing" in Hattery and Bush, editors. In press.

[9] Carlson, op. cit.

Output

Many significant refinements have occurred in the press room. They include the introduction of the belt press, automatic quality controls and automatic process control. On the other hand, these changes are evolutionary rather than revolutionary.

Revolutionary changes in output are called for by the nature of changes in the total printing-publishing technological, managerial and distributive system. The possibility of storing encyclopaedic information in computer readable form facilitates "on order" publishing.

Numbers of copies of output may range from one to a very large printing. In fact, the output may be "published" on a video screen.

Producing the output on film in a photocompositor or from a computer-output-on microfilm (COM) device, provides a flexible base to produce either film or hard copy output in size ranges and varying numbers of copies.

Management of Printing-Publishing Technology

Although the problems of the new technological systems at the operational level are indeed challenging, it is at upper levels of planning and entrepreneurship that mind-stretching and vigorous initiative are called for.

That printing-publishing is a complex process has been noted. That the steps in the process are increasingly interdependent rather than discrete has also been noted. However, the role of information science planner and manager has been ignored except inferentially.

Yet the planning-managerial implications of mechanization and automation are most significant. They call for a new total concept which should be the context of all managerial and planning action.[10]

Centralization of planning and managing is essential, even though operational decentralization is possible and for some processes supported by the new technology.

---

[10] For a case study of broad managerial concept which incorporates computer composition and other advanced technology, see the doctoral dissertation: Bracken, M. C., "An Analysis of the Evolution of the National Library of Medicine: Implications for the Development of Scientific and Technical Information Networks," The American University, Washington, D. C., 1971.

Whether centralized concern with the total process of creating text, processing text, "typesetting" or composing, "printing" and distributing is with a business organization, a professional organization, or a governmental organization does not modify its necessity.

In the United States, industrial mergers reflect these forces of unification and centralization. Many information scientists and policy officials have been slower to recognize the inevitability and significance of organizational response to printing-publishing technological change.

The potential contribution to society of balanced support for elements in the process is very great. The "information problem" can be brought under control with present and emerging technology if policy planners and managerial statesmen turn attention to it and extend support.

Research and the Information Specialist

Comparatively small investment has been made in research and development in the printing-publishing process. Most fundamentally significant innovations are "fall-outs" or "spin-offs" from research for other end objectives. The small direct investment has generally been applied to very specific objectives which have localized significance but tend to be trivial in relation to the total system.

Similar criticism can be levied against planning and support of research and development for science information.

A larger strategy for research and development and larger investment of resources is needed critically if technology is to be utilized anywhere near its potential.

Information specialists can be especially helpful in: (1) planning, developing and operating information banks; (2) assessing in advance the needs of users; and (3) providing specifications for outputs from the system. In order to carry out these functions, the information specialist must think and operate in relation to the new technology of printing-publishing as an interrelated system.

BIBLIOGRAPHY

Andersson, P. L., "Phototypesetting -- A Quiet Revolution," Datamation, December 1, 1970, pp. 22-27.

Bracken, Marilyn C., An Analysis of the Evolution of the National Library of Medicine: Implications for the Development of Scientific and Technical Information Networks. Washington, DC: The American University,1971.

CIS Newsletter. Bimonthly. Published by Composition Information Services, 6331 Hollywood Blvd., Los Angeles, CA 90028.

Composition Input: The Proceedings of the 1970 Annual Meeting. Arlington, VA: Graphic Communications Computer Association of the Printing Industries of America, Inc., 1971. 119 pp.

Doebler, Paul D., "Publication and Distribution of Information," in Carlos A. Cuadra, editor, Annual Review of Information Science and Technology, Vol. 5. Chicago: Encyclopaedia Britannica, Inc., 1970. 468 pp.

Farrell, Guy, "The Larger Meaning of Comprint 90," Book Production Industry, April 1971, pp. 30-35.

Hattery, Lowell H. and George P. Bush, editors, Automation and Electronics in Publishing. New York: Spartan Books, Inc., 1965.

Hattery, Lowell H. and George P. Bush, editors, Technological Change in Printing and Publishing Management. New York: Spartan Books, Inc., In press.

Landau, Robert M., editor, Proceedings of the ASIS Workshop on Computer Composition. Washington, DC: American Society for Information Science, 1971. 258 pp.

Lee, Richard W. and Roy W. Worral, Electronic Composition in Printing: Proceedings of a Symposium, National Bureau of Standards, June 15-16, 1967. National Bureau of Standards Special Publication 295. Washington, DC: U. S. Government Printing Office, February 1968. 128 pp.

Riley, Wayne, "Phototypesetting Today: Where is It?", Reproductions Review, April 1971, pp. 24 ff.

Seybold, John W., A Primer for Computer Composition. Arlington, VA: Graphic Communications Computer Association of Printing Industries of America, Inc., 1971. 288 pp.

Seybold, John W., The Market for Computerized Composition. Washington, DC: Printing Industries of America, Inc., 1969. 175 pp.

Session Seven - Discussions

PUBLISHING AND REPROGRAPHY
Chairman: Mr. J.H. d'Olier (France)

MR. E. OFFENBACHER (Israel): Reprography and information centers are well known subjects; however it is desirable to note what is special about the needs of developing countries in this field. In developed countries there is a steady flow of incoming requests for information to the information center. In a developing country, or as it is more adequately called, in a young country, it is necessary to stimulate the use of the information center. The information needs, therefore, are quite different from those of industrialized countries. The material to be published and distributed by the information center is also different, and so are the reprographic needs and possibilities.

Because of the rather difficult economic and administrative situation, the framework in which information centers have to work in a young country is quite different as far as the technical means are concerned. The famous slogan - what is good for America is good for you - is true only for industrialized nations. This is not because techniques developed in America, England, Russia or any of the big industrial countries haven't proved themselves, but rather because the economics of their use is quite different in a young country.

I would like to give a few examples of the sort of economic considerations I am referring to. To use micromaterial you need readers. There exist reasonably priced readers. For instance, there is an American product which sells in the US for $80 which is a very reasonable price. Translated into Israeli pounds, this is about IL 350. Add insurance, shipping costs and taxes and it comes to IL 650. You can buy approximately the same things in Israel with an Israeli pound as in the States with a dollar. So 650 Israeli pounds would mean in America $650. In the States $650 is a considerable price for a microfiche reader and it would severely limit the possibility of popularizing the microfiche system. This is a typical problem of developing countries.

Another example is the calculation to be made to find out if a certain piece of equipment is worth acquiring. The depreciation of reprographic equipment is calculated at about five years, but in practice only three years are considered realistic. After three years, there will be a new model on the market or even a new technique which will make it desirable to acquire new equipment even if the old equipment is still usable. After three years in an industrialized country, such as Germany, England, Holland or Sweden, the machine will have amortized itself. In a developing country, this is doubtful. Then there is the choice of throwing away a machine which has not yet amortized itself or of foregoing the benefit of an improved model. The question of investments and other economic considerations is thus quite different in a young country.

There is another falacy which I would like to mention here. We often hear that "any office girl can do it." Any office girl can indeed push a button, but to be sure that the machine always responds when the office girl pushes the button, we must have highly skilled technicians who can maintain and repair the machine. This again is a problem in young countries. Being situated far from centers of development in Europe and the US they don't always have experienced technicians and this makes it impossible to use the same sophisticated machinery which works beautifully in the more developed countries. On the other hand, there are prestige motivations to acquire sophisticated equipment which can't be profitably used in a developing country.

MR. M. DORON (*Israel*): There are two reasons for computer-aided typesetting of the Encyclopedia Judaica. The major reason is the shortage of trained compositors in Israel. We would have needed for the Encyclopedia around 20 trained compositors which is out of the question. You just can't find that many people on the labour market. We therefore set type on flexowriters, which require very little knowledge on the part of the typist. A compositor takes at least a year or two to train. Our typists were working after one week of training.

Rather than go through a complicated mechanical procedure, we simplified a lot of operations; for example, every time we wanted to set italic type, we simply had the typist strike a dead key which doesn't move the carriage along, in the form of a box. This punched a unique code on to the paper tape. After that for italic, she punched the letter "i". So |i| preceding a group of words or even pages would cause everything that comes after it to be set in italics.

This also simplified the preparation of the index volume. The index editors, rather than waiting until the whole text had been set and made up and then going over it and deciding which words should be in the index, took the manuscript and indexed that. In other words, if there was a certain word, for instance "Moses", which we wanted to index, the word would be marked on the manuscript in front of the word Moses $\boxed{I}$ $\boxed{1}$, The capital $\boxed{I}$ means that this is a word to appear in the index. The $\boxed{1}$ means that this is the first word in this particular article that is to be indexed. This is for later accessing. These codes did not appear in the proofs of the text itself but were in the margins for checking that the computer accepted them.

The indexer, at the same time as he marked up the manuscript, also prepared an index card. On the card he wrote "Moses", the article number, and the number of this particular word in the article that was to be indexed. He then manually sorted the cards in alphabetical order. He also edited the cards if necessary. "Moses" may not be a main item but a sub-item to some other main item. The cards were edited constantly until all the articles were included. Then they were put on paper tape on the flexowriter, and set with the word "Moses", the article number and the index number. Since we did not use automatic makeup on the text itself, we had to inform the computer exactly what article appeared on which page. The computer, as soon as it knew the article and the index number where a particular word appeared in an article, could automatically convert the article number into actual volume and page number. This is then printed out on our Photon 713, automatically made up into pages, and after corrections, is ready for printing.

We used the computer for several additional purposes, e.g. sorting of articles. We didn't set the articles in alphabetical order. The computer sorted them out and every time we were ready to close a certain article we could sort it out very simply. Anyone who has worked at a printing house would know that this would have been impossible to do without the computer.

In addition, the computer justified the lines and the typist did not have to do this at all. We used automatic hyphenation which only experienced compositors could do. Corrections were handled by updating the computer memory. Make-up of the index was also done automatically, a time consuming operation in any printing house.

All these things may not have saved money on this project, but undoubtedly saved lots of time. The index volume will appear two weeks after the last volume is printed. The whole encyclopedia will be ready by the end of the year.

PROF. L.H. HATTERY (US): My principal objective is to try to generalize for the information scientist the impact of the evolution, and to some degree, the revolution, of the technology of reprography and of printing. My thesis is that the processes of printing, of publishing, of librarianship and of information science are merging into a single complex process. It has been interesting to note that in very few meetings of information scientists has the topic of printing been identified as a major topic of concern, and it seems to me significant that at this point in time, at this Conference in Israel, that it should be identified as a major program topic. In our curriculum in information science at American University, we have had now for about three years a separate course on the automation of the printing process.

It is perhaps interesting to note that despite the general lack of attention by information scientists in professional meetings and professional journals to the technological evolution in typesetting, nevertheless, the primary impetus for the development of automatic, electronic typesetting did come from information scientists. And I think perhaps if one were to write the technology history, the information scientists at the National Library of Medicine would deserve a large amount of credit.

The papers presented by my colleagues here have been very impressive and have indicated the degree to which the advanced and sophisticated technology of reprography and printing is already operational. In the States, there are more than 1400 computer typesetting installations. More than one-third of the newspapers are being composed by computer. Although there has been a great deal of discussion for a number of years about other technological developments such as the use of optical character recognition, there has not been much operation, so we are somewhat inclined to discount the likelihood of these technologies becoming significant in the near future. I suggest that the technology and the economics of optical character recognition for the automatic input of information into composition systems, for example, is not likely to be a significant development within the next five years.

The important issue is to try to identify the useful and necessary implications of these developments for us as information scientists. One is that we must be prepared to accept lower quality output. There are examples of high quality output which suggest that it is possible with the new technologies to produce high quality material. Nevertheless, this is at the cost of more time, and a great deal of effort. There are some who are upset when they see justified right hand margins, and some when they see ragged right hand margins, because it seems that the esthetics of justifying right hand margins may not be economically justified within the practical problems of communication. I would suggest, at the risk of horrifying some of my colleagues, that hyphenless composition may be acceptable in many fields. Some of the newspapers in the States converted to hyphenless composition and very few of their readers noted this conversion. I think the same thing may be true for the readers of scientific and technical information.

Another implication of technological innovations is the expectation of highly mixed media for any complex - hard copy, microforms in various aspects, tape displays. This has grave implications for our forward planning.

Let me suggest several of the specific problems that we ought to be taking into consideration in forward planning. There has been some reference to the matter of building to hold new collections of books. One of my battles is against library architects. In the information sciences we who need to build a new building must depend on architects who have been trained to build a building to hold books only. Even within my own university, the principal concern and goal of our university librarian has been to build a large new library to hold more volumes. It seems that the investment ought not to be in physical facilities for traditional hard copies but rather for the mixed media which are already available and which can be made available to substitute for hard copies.

A second concern in the area of planning is the establishment of networks. Those who are engaged in planning networks beginning with UNISIST through to component networks, need to have greater input from information engineers to devise standards for media mixes which are optimal for large networks. It seems to me that because most of the documentation that relates to network planning does not take technology adequately into account, plans will prove obsolete within five to ten years.

Finally I suggest that the developments that I have been referring to in technology and their applications to major forward planning in terms of facilities, equipment, personnel and networks, call for a substantial degree of centralization of planning and support. In the US and other countries, governments and professional associations have not given adequate attention to financial support for the kind of planning and preparations necessary for the effective and optimal use of the new technology.

DR. M. CREMER *(Germany)*: I would like to mention two possible applications of microfiche. The first is in the distribution of doctoral theses. In my country, we have the problem that, on the one hand, the students and the faculty do not want to shoulder the financial burden of printing theses in conventional form. On the other hand, the university libraries are very keen to get copies of theses and dissertations for exchange purposes. We started to develop a national program to put doctoral theses on microfiche and I hope to convince my library colleagues to accept this program.

The second application relates to archives and office records. Large offices like governmental departments, archives, big scientific organizations, etc. put their records on microform not only to save space, but also to have access to them through a combination of microform and computer search, thus making the records of the office or of the archives available as a kind of integrated information service.

MR. OFFENBACHER: *To Dr. Cremer*. I have found that new Ph.D.'s who just presented their theses are rather reluctant to make them available before they have appeared in the press. Only a few copies are required for the university authorities. The question is if microfiche is economical under these circumstances, since I doubt that professors would be prepared to read the microfiche in order to give their opinion about the thesis.

MR. D'OLIER: In France we have a program for all French speaking countries, many of which are developing countries, so the question of price of micro-documents is of paramount importance to us.

MR. P. TURA *(Israel)*: *To Mr. Thiele and Mr. Offenbacher*. Because a small country can't afford to buy a lot of sophisticated and expensive equipment, the manufacturers and designers of this equipment should get to work. Microequipment for the viewing of microfilm and microfiche could be produced at a cost not exceeding by much that of a pair of good spectacles. I have seen some readers for viewing film slides for tourists and possibly this or a stereoscopic viewer might be adapted for reading microfilm, at low cost. This would allow every scientist to buy a cheap microfilm-microfiche reader which would operate in daylight or on batteries. Microfilm-microfiche can also be viewed with a slide projecter.

MR. G. THIELE (Germany): *To Mr. Tura.* A manufacturer can make cheaper readers if he can make them in large quantities. This is only possible if he is sure of a large market. If there are various microfiche sizes you have to have various types of equipment. In Frankfurt we have an automatic enlarger for microfiche.

MR. OFFENBACHER: *To Mr. Tura.* I am afraid what works well for a tourist looking at slides wouldn't work as a reader for microcopies. There are too many details to be read which we don't look for in a snapshot.

MR. U. PUMPIAN (Israel): *To Mr. Thiele.* Could you give us some more details about the plan to publish abstract periodicals? Is it done in co-operation with publishers? It would be useful to shorten the period between the time data is collected and the time it is published.

MR. THIELE: *To Mr. Pumpian.* With the periodicals there is some trouble, mainly the problem of copyright. When there will be some agreement on copyright I think there will be agreement between the producers and the users.

MR. L. PAPIER (US): Research is being carried out in the United States on the acceptability and effectiveness of microfiche. One study has been completed at the University of Denver on the use of microfiche. We have a big project with the Am. Assoc. of Junior Colleges in a learning situation with the whole course produced on microfiche, highly controlled where variables are determined. These variables are things like whether it is data or information, positive or negative film, type of reader, etc. I just want to bring out that we do have a little more hard data on microfiche, microform than we had in the past.

MRS. S. WEIL (Israel): Users of microfilm or microfiche use microform collections for eliminating not needed material. If they find a reference or if the SDI service brings to their attention a document which exists only on microfiche they come to our library and read only the abstract or the main points. If the document is of interest, they ask for a full copy. The microforms are more for the archives than for the user. I would like to know if my colleagues in other countries have the same experience or if their users are really reading the microfiche and using them at home with a reader.

MR. Y. SHALEV (Israel): In the documentation unit at the Ministry of Defence we recently distributed some 35 readers to our users. The price of each reader is \$60 - \$70 and is paid off in a year if only 35 reports of MPIS are read per annum, as the microcopy of the report costs only 95 cents. Another economic point which should also be considered is that storage space in the Tel Aviv area is IL 1,000 per sq./m. per year.

MR. OFFENBACHER: *To Mrs. Weil.* You are quite right. Microcopies were never intended to be read but only to be scanned. When it becomes apparent that a whole publication is to be read, the text should be enlarged.

MR. THIELE: *To Mrs. Weil.* I agree that it is impossible to read a whole microfiche of 60 pages. We can make copies in about 3 minutes, and the price for these 60 pages is 12 Marks.

If a publisher produces a printed periodical and a microfiche, in Germany he receives the same price for both. In the case of reports, there is only the microfiche as the manuscript is used only for filming.

MRS. WEIL: *To Mr. Thiele.* It seems very important to have the possibility to acquire journals both in the form of microfiche and the full size copy. For browsing, readers need the latter, whereas for back searching the microfiche is very suitable. If we had a combination price that would allow us to buy both at a reduced price, we would save storage space as we could discard full size back volumes.

MR. *U. BLOCH (Israel):* I don't see any reason why someone who wants to buy the full sized journal and the microfiche should pay the editor and the author twice. As Mrs. Weil said, we need the hard copy because people want to browse and we need the microfiche to save storage space. Why should the microfiche not be available at the actual cost of preparing it which would come to something like ten per cent above the price of the hard copy. This would enable librarians to start working in a new direction.

To buy the film later is expensive and if the microfilm is prepared not from the original its quality is entirely different. If there are any representatives of publishers present, they should take this discussion into consideration and think about offering the microforms at a reasonable price.

In addition, I'd like to know whether the speakers have any idea of what will be the next step in the development of faster typesetting machines. Perhaps on-line displays photographed directly to the film, and then shipped to the printer might be the next step.

MR. P. SIMMONS (UK): To Mr. Bloch. New composing machines come out every day. Speeds of something like 10,000 characters a second have been recorded. People have problems with powerful computers keeping up with them. I can give you the names of these machines but they are probably meaningless. Speed at the composing end of the system is not the problem. Time span from the acquisition of the material to its actual publication is very much longer than the two or three days of work on the 713. This could obviously be reduced to a few hours but with a time schedule of two or three months, it makes little difference.

DR. S. SCHWARZ (Sweden): To Mr. Zuchel. The question of transformation of text to magnetic tape is one of time saving. You have a system where four-punches represent one character. The Japanese have developed a system which enables them to represent about 2,500 characters with two-key punching which is done not sequentially but simultaneously. This punching seems to go with great speed. Have you a comparative economy of these different systems?

MR. SIMMONS: To Dr. Schwarz. We have a similar problem with some 700 different characters we need to encode. We have adopted the technique of taking one character followed by a shift number of which there are six, and following that with a specific character. For example, we would take a character which corresponds directly with the paragraph sign and follow it with "1" for the Greek alphabet, and "a" thereafter would indicate alpha and so on.

MR. H. ZUCHEL (Germany): To Dr. Schwarz. A system of prototypes which we use is open ended; the only thing you have to do is cover the frames of the composition machines with the systematic tables for Japanese characters, for example. You can put in these tables what you want; all characters are possible.

DR. SCHWARZ: To Mr. Zuchel. It was a question of timing. Does it go as quickly as punching two keys simultaneously?

MR. ZUCHEL: To Dr. Schwarz. Yes, it is a question of timing, but timing is not interesting in this case. The problem is to record all the data without loss of information. We have solved this with this system. You do need more time for four representatives of one character.

MR. L. ELPERN (Israel): To Mr. Doron. Are you considering the possibilities of making any form of micro-reproduction of the encyclopedia? Is the body of knowledge accumulated in this encyclopedia going to be updated in some way? Are you considering any kind of secondary publications from the accumulated knowledge?

MR. DORON: To Mr. Elpern. No microphotography is anticipated as yet, but we will consider your idea. As for future updating, that of course is one of the prime reasons that we use the computer. Future editions of the Encyclopedia will be much easier and quicker to update than would have been ordinarily possible. Updating within a few years is anticipated. The third point - using the accumulated knowledge in other forms, such as separate books or other publishing ventures, is also being planned. We plan many side products to the Encyclopedia. A huge amount of knowledge has been gathered and since it is computerized, it is readily accessible.

SESSION EIGHT

commercially available services

Authors of Papers

THE INTERNATIONAL NETWORK OF ISI®-LINKED NATIONAL INFORMATION CENTERS

Eugene Garfield, Ph.D. and Morton V. Malin, Ph.D.
Institute for Scientific Information, Philadelphia, Pa., U.S.A.

SYNOPSIS

During the past six years The Institute for Scientific Information has, through private investment, constructed a multidisciplinary machine-language data base covering nearly two million source articles. This file is increasing at present by approximately 400,000 items per year or 8,000 per week. Use of ISI's information services offers a method of establishing an effective national information system at a cost acceptable to smaller and developing nations.

Customarily, discussions of information networks are concerned with the concept of interrelating the activities of information centers on a national or international scale. Only slight, if any, attention is given to the already existing international network for scientific information exchange which is the scientific community itself. Scientists have disseminated information and communicated with one another using a variety of methods, but primarily through the published literature (journals), correspondence, scientific meetings and more recently telephone and other modern media of communication, for several hundred years.

It is understandable that discussions of information networks do not become involved with these first order networks but concern themselves with, what I shall call, second and third order networks. The second order networks are those intended to establish a system of information interchange on an international level, and the third order networks are those which involve systems for linking national information centers. The concern with second and third order networks is understandable because the processors and disseminators of information are faced with the problem of managing the information flow that results from the research and scholarly activities of the scientific community. The concept of network systems, international or national, reflects the desire to make the output of the scientific community more readily available to a greater number of potential users of information, and to do this more efficiently by establishing methods which will reduce duplication of effort. These objectives are important because all of us are aware of the duplication of effort that exists and of the high price that is paid as a result.

Too often, however, the demand for efficiency translates itself not into making better use of that which already exists or improving the operation of existing systems, but into developing new and different methods and systems.

VIII Commercially Available Services

ISI's approach to establishment of a network of internationally-linked information centers has been to develop better methods for dealing with the communication system that already exists among scientists and to expand and nourish it. This is done not only by computerized services, which is the main focus of my talk, but by a variety of services, both manual and machine oriented, that play a significant part in ISI's role as a processor and disseminator of information at all levels -- individual, institutional, national, and international.

Before considering the relationship of ISI to the information centers located in more than a half dozen countries in which ISI tapes are used, it is important to describe ISI's approach to providing the totality of its services to individual countries in their efforts to establish national information centers which can link up to the ISI network.

ISI's services can be characterized both functionally and by subject area. The first category includes:

1. Current Awareness Services
2. Retrospective Searching Service
3. Selective Dissemination of Information Service
4. Library Service.

Although the subject areas covered in ISI services include every field of scholarship, I shall delineate these areas broadly as follows:

1. Life Sciences
2. Physical Sciences
3. Chemical Sciences
4. Agricultural Sciences
5. Engineering and Technology
6. Behavioral and Social Sciences
7. Education.

I mention these subject areas only to indicate the broad coverage available through ISI's services. More important, I believe, is the description of the types of services offered as categorized by function, although I shall at a later point talk in some detail about ISI's chemical information service.

Using as input approximately 5000 journals from almost every country in the world, current awareness service is provided through the medium of five Current Contents® editions. These are:

1. Current Contents/Life Sciences
2. Current Contents/Physical & Chemical Sciences
3. Current Contents/Agricultural, Food, & Veterinary Sciences
4. Current Contents/Engineering & Technology
5. Current Contents/Behavioral, Social, & Educational Sciences

The purpose of Current Contents is to provide prompt current awareness of research results published in the leading journals of the world in the broad areas shown above. Each week the contents pages from the journal issues received at ISI are reproduced in one of the five Current Contents editions. Each Current Contents also includes an author address directory which readers can use to send to authors for reprints of articles they have found through Current Contents. It has been estimated that five million reprint requests per year result from use of Current Contents. An important feature of the current awareness service is the fact that ISI has established relationships with publishers throughout the world which enable us to obtain journals quickly; the contents pages of many journals are obtained in advance of publication.

Retrospective searching capabilities are provided by ISI through the Science Citation Index® (SCI®) and Permuterm® Subject Index (PSI) and ISI's Search Service. The SCI and PSI are comprehensive multidisciplinary indexes which provide a variety of unique methods for retrospective searching of the scientific, agricultural, technological, and biomedical literature. The SCI provides access to the literature by author, citation, or organization, and the PSI is a natural-language subject index which complements the SCI, providing access to the literature by words taken from the titles of current articles processed into the SCI data base.

The SCI data base is derived from a core of approximately 2500 of the 5000 journals received at ISI. Each journal issue is indexed into the system from cover to cover and all substantive items are processed. The SCI and PSI are published quarterly with annual cumulations. In 1971, the data base will consist of approximately 400,000 current articles and 4,000,000 citations. This year will also see the publication of a cumulative SCI covering the years 1965-1969. The total SCI file now includes over 1,500,000 source items and 20,000,000 cited references.

Retrospective searching is also provided through the ISI Search Service which is used by many subscribers when they require special literature searches to be performed.

SDI is available from ISI either through ASCA®IV, the Automatic Subject Citation Alert service operating at ISI's headquarters in Philadelphia, or through lease of the Science Citation Index tapes which can be used by organizations to provide SDI to their staff. The data base used for ASCA, or available from ISI tapes, is the same one which is used for the SCI. SDI services, however, are provided weekly through subscription to ASCA or, if tapes are used, these are also supplied weekly to the lessee. The ASCA service is unique in the fact that citation questions, as well as words, can be used to develop the user profiles. Tape users also have this feature available to them if they wish to obtain the citation tapes as well as the source tapes, and very important is the fact that the ASCA software can be bought or leased from ISI.

Among the functional service areas mentioned previously was ISI's library service. This service called OATS®, Original Article Tear Sheet service, was designed to provide users of ISI's system with the means of obtaining hard copy of any article which their use of any of the current awareness, retrospective searching or SDI services uncovered, when such articles are not readily available through regular library channels. In fact, in many cases the OATS service is used by libraries and individuals even if the journals are available in their libraries because of the convenience and the speed with which they receive tear sheets and because, on a cost benefit basis, the service is relatively inexpensive. Some ASCA customers have requested that articles reported on their printouts be sent to them automatically; we call this service ASCAMATIC.

I have only, in the preceding, described very briefly the four major service functions performed by ISI which ISI feels are solutions to the four basic aspects of the problem of disseminating scientific and technical information. These solutions are keeping users of information current on scientific and technological developments: Current Contents, keeping them informed on a selective basis of developments having direct bearing on their work; ASCA and Tapes, enabling them to learn quickly what has been published in the past; SCI, retrospective searching capability; OATS, enabling them to retrieve documents quickly and efficiently.

I stated previously that ISI has a specialized chemical information service which I shall describe briefly before commenting upon the ISI-linked international information network.

ISI's chemical information services group consists of the publication Current Abstracts of Chemistry and Index Chemicus™ (CAC&IC™), the Index Chemicus Registry System®, and the Chemical Substructure Index™. Current Abstracts of Chemistry and Index Chemicus is a weekly abstract journal of the chemical literature which places special emphasis on reporting new compounds and reactions. In addition to the traditional abstract, structural diagrams and reaction flow diagrams are also provided. Indexes to the weekly issues are published monthly and cumulated annually. Current Abstracts of Chemistry and Index Chemicus is the source for preparing the Index Chemicus Registry System. The compounds reported in CAC&IC are encoded into Wiswesser Line Notations, which permits preparation of magnetic tapes that are computer searchable. Searching can be by compound family, substructure, biological activity, authors, journals or index terms since the tapes contain not only the encoded compounds, but also bibliographic information for the article in which the compounds were reported and information regarding the other searchable items mentioned. Tapes are provided monthly and programs are available for use both for retrospective and current awareness searches of the file.

The Chemical Substructure Index is derived from the ICRS® and is a published permuted listing of the compounds encoded into WLN. This Index provides the capability of doing substructure searches for the most recently reported chemical compounds by manual means. It is published on a monthly basis and will be cumulated annually.

My purpose in providing this brief overview of ISI services is to lay the basis for describing how ISI serves as a mechanism for linking information centers and explaining how ISI's services can be obtained on a national level to help in development of national information systems.

ISI is a supplier of its SCI data base to national centers in five countries for SDI use and also supplies these tapes to approximately half a dozen private organizations which use them for internal SDI services. The ICRS tapes are presently supplied to a number of private chemical firms but are also available to national centers.

In making these tapes available on a lease basis, ISI has, in fact, already aided in the establishment of an international network of information centers in which ISI, as the processor of information produced on an international scale, services as a distributor of reprocessed information to an international clientele. The present system is not a formalized one with rules, by-laws, standards, etc. But, there is no reason why the process cannot become more formal. In fact, since ISI is prepared not only to lease its tapes but also to lease or sell its whole ASCA software system and to help in its installation and guarantee its operation to any country, many of the obstacles which inhibit establishment of networks are eliminated. There exists at this very moment the capability for every country represented at this meeting to have, without the need for excessive development costs, a national SDI system, simply by obtaining from ISI the rights to the ASCA system.

Even more important, the whole ISI system consisting of the four components, current awareness, SDI, retrospective searching capability and library service, is also available because ISI is prepared and willing to cooperate with national information systems to supply all its services and publications to a central authority for redistribution to the scientific, engineering, technological, educational, and managerial communities in each individual country.

Such an arrangement has already been negotiated with the Ministry of Education in Spain, negotiations are now in progress in several Latin American countries for obtaining all ISI's services on the national level. Negotiations have been completed with a representative of the Japanese Government for use of the ASCA and ICRS system on a national level in that country. I am also sure you know that ISI tapes are used by the National Research Council in Canada, the Royal Technical Institute in Sweden, and by COSTI here in Israel.

In its arrangements for service on a national scale as, for example Spain, ISI provides to the central authority or national system bulk shipments of the five editions of Current Contents and Current Abstracts of Chemistry and Index Chemicus, sets of the SCI for placement in universities, research institutes and other libraries, provides OATS service through the national center, and Search Service on demand.

We aid in the establishment of a national SDI system through a phasing-in process by supplying ASCA from Philadelphia until the computerized system is installed locally and by providing technical and educational service. These services are designed to guarantee the operation of the system and to train both the administrators and users of the system in the techniques of profile preparation and refinement, and in managing the system. I wish to stress very strongly that ISI takes its responsibilities in such cooperation very seriously because we insist on providing training to local personnel on the operations, administration and use of our system. The management problems of building a system are well known to all of you and the literature is full of articles reporting on these problems; ISI's objective is to eliminate them.

ISI stresses that it will provide complete administrative and management support in connection with the supply of its services. This support includes advising on the type of organizational structure that should be established for the system and in providing training for clerical and technical personnel who will be responsible for operation of the system. This training will include instruction in record keeping, acquisitions procedures, and methods for establishing distribution procedures for the published and computer-generated services, and for establishing financial and administrative procedures to insure efficient operation of the system.

In regard to the ASCA system, ISI provides as part of the software lease full documentation of the programs. Its computer systems personnel will help install the programs at the local computer center, and train operators and systems and programming people in its use. ISI will also provide training on profile preparation and profile input processing at the local center and to personnel sent to Philadelphia. Finally, I should stress that ISI is prepared, as part of its training responsibility, to develop instructional materials on all its services in the local language.

I shall conclude my talk by saying that it is universally recognized that the future growth of any country in the world today depends on developing a strong educational and research and development capability. Development of this capability is a major requisite for establishing the base for future economic growth and solving the pressing social problems faced by all nations. A major contributor to the solution of these problems will be properly controlled and disseminated information to the individuals who are engaged in working toward the solutions. The cost, to any nation, of solving the information problem by itself is very high and in many cases actually prohibitive. The answer, consequently, is to take advantage of resources already available as well as to develop systems to meet specific individual needs. Recognizably, no country wishes to remain completely dependent on outside resources indefinitely, but also recognizable is the fact that economic necessity and good management require that the development of national information systems be done in the most efficient manner possible. Exploiting available resources which are cheaper, costwise, than development of new similar resources is an important way to keep development costs low.

I suggest that effective national and international systems and networks can be developed by linking into available systems such as ISI's.

BIBLIOGRAPHY

1. Garfield, E. "A Unified Index to Science." in Proceedings of the International Conference on Scientific Information, Washington, 1958. (Washington, National Academy of Sciences-National Research Council, 1959), Vol. I, pp. 461-74.

2. Garfield, E. ISI services in the design of small user systems. Journal of Chemical Documentation 6(3):164-8, 1966.

3. Sher, I.H., Garfield, E. & Elias, A.W. Control and elimination of errors in ISI services. Journal of Chemical Documentation 6(3): 132-135, 1966.

4. Garfield, E. & Sher, I.H. ISI's experiences with ASCA® -- a selective dissemination system. Journal of Chemical Documentation 7(3):147-153, 1967.

5. Malin, M.V. & Levy, W.A. "ISI's SDI System." in Impact of Mechanization on Libraries and Information Centers. Fifth Annual National Colloquium on Information Retrieval, ed. by James W. Ramey (Philadelphia, Information Interscience Inc., 1968) pp. 51-61.

6. Malin, M.V. The Science Citation Index®, a new concept in indexing. Library Trends 16(3):374-387, 1968.

7. Garfield, E. "World Brain or Memex? Mechanical and Intellectual Requirements for Universal Bibliographic Control." in The Foundations of Access to Knowledge, ed. by E.B. Montgomery (Syracuse, Syracuse University Press, 1968) pp. 169-196.

8. Garfield, E. Permuterm® Subject Index -- the primordial dictionary of science. Current Contents®/Life Sciences 12(22):6, 1969.

9. Garfield, E. The Who and Why of ISI. Current Contents®/Life Sciences 12(12):5-9, 1969.

10. Garfield, E. "Citation Indexing, Historio-Bibliography, and the Sociology of Science." in Proceedings of the Third International Congress of Medical Librarianship, Amsterdam, 5-9 May 1969, ed. by K.E. Davis & W.D. Sweeney. (Amsterdam, Excerpta Medica, 1970) pp. 187-204.

11. Garfield, E. "The Role of Man and Machine in an International Selective Dissemination of Information System." Paper presented at the International Congress of Documentation of the International Federation of Documentation, Buenos Aires, 1970.

12. Weinstock, M. "Citation Indexes." in Encyclopedia of Library and Information Science (New York, Marcel Dekker, 1971) Vol. 5, pp. 16-40.

CAN/SDI PLUS TWO
CANADA'S NATIONAL SDI SERVICE
FOR SCIENCE AND TECHNOLOGY

Jack E. Brown
Director, National Science Library, Ottawa, Canada

SYNOPSIS

The National Science Library's national SDI Service, after two years of operation, is proving to be a successful and valuable service. Employing six data bases covering 900,000 papers per year, it provided in 1970, 52,000 tailor-made bibliographies to 1400 users. Steps in the development of the service are summarized, strengths and weaknesses of the system discussed, and plans for future development outlined.

The Canadian Selective Dissemination of Information Service (CAN/SDI), a national computerized information retrieval system, for Canadian scientists and technologists, has now been fully operational for two years and three months. During this period five papers have been published in the national and international literature dealing with specific aspects of the service - e.g. systems design and programming, training of users, and user reaction to the service[4,5,7,8,9]. This, the sixth paper in the series, is an attempt to provide an overview of the CAN/SDI system with particular emphasis on the strengths and weaknesses of the service from the point of view of the providers and users, and a look at plans for future development.

For those of you who have read these earlier papers we apologize for what may appear to be inexcusable and useless repetition. For those who are not familiar with the service, we apologize for the omission of technical details which are certainly important but which cannot be dealt with adequately in a paper of this nature. For both groups, particularly those who are experimenting with or considering a similar service, we trust this summation of our experience will prove interesting and useful.

VIII Commercially Available Services

The National Science Library of Canada

The CAN/SDI service is one of many national information services provided by the National Science Library (NSL) and to see this program in proper perspective, it is necessary to briefly review the role of the NSL as Canada's major agency responsible for the dissemination of scientific and technical information (STI).

The NSL, a Division of the National Research Council of Canada (NRC), is not a library in the conventional sense of the word but rather an information transferral agency. Its total activities are designed to supplement and complement local information services and to ensure that Canadian scientists, technologists, industrialists and managers, have ready access to publications and information required in their day to day work. The NSL's resources, now totalling some 800,000 volumes, and its information services, have been developed in close cooperation with all major libraries in Canada. Through these and related cooperative measures, and through the utilization of Telex linkage with world-wide information services, the NSL serves as the focal point of a national scientific and technological information network.

A detailed description of the nature and extent of the NSL's resources and services may be found in the NSL's Annual Report and Guide to Resources and Services 1969-70[6], and need not be repeated here. However, to emphasize the rather unique role of the NSL, as compared with the information services of similar agencies in other countries, two points should be noted:

1. The NSL's information retrieval and dissemination services and the essential backup resources form an integrated administrative unit.
2. The NSL is the major unit of a national decentralized scientific and technological information system.

Development of a National STI System

The NSL had its beginnings in 1925 as the library serving the NRC of Canada. Its change from that of a library serving a group of federal laboratories to that of a national information centre has come about through a long series of government decisions and directives - the change has been evolutionary not revolutionary. One of the most important of these decisions was made in December 1969 when, as the result of a series of government studies of science and science information policies[1,2], the NRC was named the coordinating body responsible for the further development of a national STI network.

This action recognizes the effectiveness of a decentralized information system in a country the size of Canada as opposed to a highly centralized system. It also represents a crucial point in the development of Canadian information policy, for the government directive contains clauses intended to ensure that a national system for STI, which is the responsibility of the NRC, and a national system for information in the humanities, social sciences and the arts, which is the responsibility of the National Library, are completely compatible and developed in parallel. A full account of this development may be found in a paper entitled "A National STI System for Canada"[10].

CAN/SDI - Basic Principles

The NSL is continually experimenting with and developing new techniques to facilitate the storage, retrieval and dissemination of information. The CAN/SDI system is one of the more successful of these techniques and is regarded as an extension of the Library's Reference and Research Services and as another tool to assist researchers obtain maximum use of pertinent literature.

The CAN/SDI service was inaugurated as a national service in April 1969, after three years of testing and experimentation using NRC and other Ottawa scientists as guinea pigs. During this experimental period three major problems were identified as requiring solution before a national service could be provided.

1. The incompatibility of available data bases or tape services from the point of view of file formats.
2. The development of interest profiles for users remote from Ottawa.
3. The training of potential users of the system.

Development of a Standard File Format

The solution of the first of these problems, incompatibility of data bases, was the one requiring urgent attention and the one which provided most difficulties. It was obvious that anytime we wished to add a new data base to the system, a new search and print phase had to be written - a costly and difficult undertaking. Accordingly, after much experimentation, a standard file format, based on the MARC format, was designed. The steps leading to the development and adoption of this format are described in a paper by J. Heilik[9]. All data bases used in CAN/SDI are converted to this standard format thus requiring only one search program for the data bases and one program to print the results of this common search. The MARC format is extremely flexible and can handle with ease almost any kind of information.

With the development of this standard format new data bases were rapidly added and at present five tape services are being employed - Chemical Titles, Chemical Abstracts Condensates, ISI Source and Citation, INSPEC and Biological Abstracts. A sixth data base, MEDLARS, is also being used to provide both current-awareness and retrospective searches, but the programs used for query formulation and computer searches are those developed by the U.S. National Library of Medicine. At present search requests are formulated at the NSL, but until the MEDLARS II programs are available for use with the NSL's computer facilities, the tapes are being processed at the Texas Medical Center in Houston.

These data bases provide users of the service with access to the contents of approximately 9000 journals or 900,000 papers per year in all fields of science and technology. All cited journals are included in the 16,000 periodicals received by the NSL and photocopies of cited papers, not available locally, are provided upon request.

Through the use of an IBM 360 Model 65 computer, interest-profiles are matched against authors, titles, journals, key words and/or descriptors as stored on the tapes mentioned above. Each subscriber receives weekly or biweekly, depending on the tape services which he uses, a computer print-out of references to papers covering his specific fields of interest. The print-out is on sheets perforated to produce cards the size of IBM key-punch cards. Duplicate prints containing questions regarding the relevancy of the citations facilitate feedback from users to the NSL. A sample of a typical print-out with a key to the format is shown in Figure 1.

Solution of the second problem, the development of interest profiles for users remote from Ottawa, while presenting fewer technical difficulties, has proved equally costly and time consuming.

Training of Search Editors and Users

Again, during the experimental period, it soon became evident that while the searching of data bases could be centralized, the creation of interest profiles must be decentralized. Our present experience indicates that at least ten data bases, providing an SDI service to 10,000 users, can be readily processed centrally using appropriate computer facilities. On the other hand, it is equally evident that effective interest profiles are more difficult to construct via telephone conversations or correspondence, and that it is essential to arrange for face-to-face conversations between the subscriber and someone fully trained in the art of profile construction. This is indeed a major problem in a country where distances of up to 3000 miles can separate potential subscribers from the focal point of this SDI service. The problem was solved in two ways.

Seminars were established to train Search Editors in the art of profile construction[7]. These two-day seminars are presented at the NSL in Ottawa, and initially were held every two weeks. Now that the preliminary rush has subsided, the seminars are held as required, usually once a month. To facilitate training, attendance is usually limited to ten participants. During this period the pupils are thoroughly acquainted with the main features of the SDI program and are given practice in the actual techniques of profile construction.

Search Editors are responsible for assisting SDI users in their respective organizations and constitute a direct link between the NSL and subscribers in all parts of Canada. They are also responsible for submitting properly constructed profiles to the NSL, for updating and sharpening profiles once they are operational, and for scanning and distributing the printed bibliographies to the users they have assisted.

CAN/SDI - Print-out Format

```
1 X RAY, FLUORESC, SPECTR, METAL

  SHONO T      OSAKA UNIV., SUITA, JAPAN).
  SHINRA K

  DETERMINATION OF METALS IN AIR-BORNE DUST BY X-RAY FLUORESCENC
2 E SPECTROMETRY USING A FILTER PAPER TECHNIQUE.

  BUNSEKI KAGAKU
  VOLUME 0018, ISSUE 0008, YEAR 1969, PAGE 1032-4

3 MANGANESE ** DETN ** AIRBORNE ** DUSTS ** METALS ** ZINC ** IR
  ON ** TITANIUM **

  AN 0060200 P 8895 EN 01 TW 000 WT 000 S C0272 TP ARTC L
          4        5      6     7      8       9      10    11
```

(1)	Profile words that caused retrieval	
(2)	Full reference	
(3)	Keywords or abstract (if available)	
(4)	Accession or abstract number	(AN)
(5)	Profile number	(P)
(6)	Expression number	(EN)
(7)	Threshold weight	(TW)
(8)	Weight total	(WT)
(9)	Tape service, issue, year or volume	(S)
(10)	Type of publication	(TP)
(11)	Language (if available)	(L)

Figure 1

To supplement the training sessions held in Ottawa, the NSL conducts seminars in other parts of Canada where there is a concentration of potential subscribers, as for example at large universities and industrial parks.

As of April 1971, 250 Search Editors have been trained. Of this number 30% came from federal and provincial government agencies, 40% from universities, 20% from industry and 10% from hospitals, public libraries, provincial research councils, etc.

The training of Search Editors and of users has also been facilitated through the preparation of a "Profile Design Manual"[3] which is distributed without charge to Search Editors and SDI users. The Manual, now in its second edition, is periodically reworked and updated to incorporate refinements in the system and to clarify instructions. Some research workers have found the Manual sufficient to enable them to construct effective profiles without the intermediary of a Search Editor. The majority of subscribers, however, prefer to work directly through a trained Search Editor.

Costs and Subscription Fees

Now a few words about costs. Until recently subscribers to the CAN/SDI service were charged $100 per year for an interest profile of up to 60 terms and the use of one data base. An extra $15 per year was charged for each additional data base used. Two major factors were considered in establishing this fee. It was designed to recover the major costs of providing the SDI service, and to produce a fee which would discourage the diletante, but not deter the serious research worker. The figure, while initially little more than an educated guess, accomplished its purpose for within two months after announcement of the service, 210 subscribers had signed contracts, and this for an experimental service which, as far as users were concerned, had not proved its worth.

After two years of operation we are now serving approximately 800 subscribers requiring a total of more than 3500 individual computer queries, and resulting in the preparation each week of over 1000 personalized bibliographies. Because of group profiles and profiles containing multiple questions, the program actually serves 1400 individuals representing 7000 questions or subject interests. During the past year 52,000 tailor-made bibliographies were supplied to SDI subscribers.

This growth in the use of the CAN/SDI service has come about primarily through word of mouth between satisfied subscribers. Except for brief announcements in NSL Newsletters and articles in national journals, no attempt has been made at large scale publicity measures. Staff shortages and inadequate budgets made it impossible to cope with a faster rate of growth in the number of subscribers. However, NSL staffs and budgets have been increased in keeping with the wider responsibilities now assigned to the NSL, and an extensive publicity campaign is being prepared. With such measures it is anticipated that subscriptions will double within the next 12-18 months.

The CAN/SDI program is carried out at the NSL by a staff of eight - a director, assistant director, one systems analyst, two search editors, two clerical assistants and one key-punch operator. This number does not include staff support provided by the NRC Computation Centre.

After two years of service the CAN/SDI program has proved to be a valuable information retrieval service. During this time the NSL has also had an opportunity to determine the actual costs of providing this service. Based on this new information and effective April 1, 1971, a new pricing schedule was implemented to reflect actual use, computer time, staff salaries, leasing of tapes and mailing charges. The annual subscription concept has been replaced by a system of charging per tape searched. A minimum fee of $40, which provides for a profile of 60 terms is now being charged. At the end of each 12 month period the subscriber is invoiced for an additional amount based on his use of the data bases available. Under this system, and with a profile not exceeding 60 terms, users of ISI Source tapes and CA Condensates will be charged $117 per year. Users of Chemical Titles, INSPEC, Biological Abstracts and MEDLARS will be charged $45, $78, $81 and $60 respectively. The new schedule will thus result in increased costs for some subscribers but for others there will be a considerable reduction.

Feedback from users, via the duplicate citation print-out, has clearly indicated the usefulness of the CAN/SDI service and the general satisfaction of subscribers. The value of the service was confirmed when, after one year of operation, 80% of the original users renewed their subscriptions. In no case was there a cancellation because of dissatisfaction with the service. Cancellations were due primarily to changes in occupation or responsibility, lack of the necessary fee, or unsuitability of the data base.

User Evaluation of CAN/SDI

Because of the experimental nature of the program, the high investment in tape services, staff and computer time, and in spite of favourable response, it was decided that a wider sampling of user reaction was required in order to plan future developments. In June 1970 a questionnaire was mailed to 604 users of which 406 replied. The questionnaire sought expressions of opinions regarding completeness of subject coverage, additional tape services, deficiencies in the system, multiple use of computer produced bibliographies, and usefulness of the service[8].

Coverage of chemistry, physics, electrotechnology and mathematics was judged good, but there was a strong demand for better coverage of the biological sciences. The NSL responded by adding Biological Abstracts and BA Previews to its list of data bases.

VIII Commercially Available Services

There was an overwhelming demand for a retrospective searching service, which up to this time is available only through the MEDLARS tapes. Studies are now under way to determine whether the CAN/SDI service, which is essentially a current-awareness service, should be expanded to allow for retrospective searches, or whether a complementary national service to deal with back-file queries should be established.

A major criticism of the service centred on the delays experienced in acquiring hard-copies of cited papers. This comment was particularly disconcerting since the NSL, in initiating the SDI service, had guaranteed that all source literature covered by the various data bases would be available either in its own collections or in those of local libraries and information centres. Clearly there is a continuing need to strengthen the resources of local libraries. Furthermore the communications processes whereby information is transferred from the NSL to the ultimate user, must be improved. With the establishment within the NRC of an Advisory Board for Scientific and Technological Information and increased support of NSL resources and services, the prospect of correcting these deficiencies is most encouraging.

The questionnaire raised the perennial problem of relevance. Some users of the CAN/SDI service are satisfied with a 10 - 15% relevancy, while others want and claim to have achieved 95% relevancy. In any case the comments indicated that only the user can and should decide as to what is relevant, marginally relevant, or irrelevant. The comments also underlined the importance of developing a satisfactory procedure for preparing individual profiles and for continual adjustment through feedback.

The question regarding usefulness of the service produced some startling information - namely that those users who responded to this question had achieved a combined saving of 39,000 man-hours per year gathering interest related information through use of the CAN/SDI service. A surprising number of users also indicated that, because of their use of CAN/SDI, they had given up conventional search procedures.

The survey confirmed that the CAN/SDI is an effective current-awareness service and that it is becoming an important element in the Canadian national system for the dissemination of STI. It has achieved considerable success in reducing the time spent by researchers in monitoring the current literature while at the same time providing increasingly reliable and extensive coverage for industry, university and government.

Plans for the Future

The ever expanding need for information by science, government and industry, and the rapid improvement in techniques for processing information, makes it extremely difficult and perhaps unwise to make plans for future action, unless the plans can be readily changed. In these matters the NSL's crystal ball is no clearer than ones used by other countries which are also struggling to improve the dissemination of information.

With these reservations in mind, and speaking only of the CAN/SDI Project, the NSL does plan to extend and improve its SDI services along the following lines:

1. Extensive publicity to acquaint more Canadian scientists, technologists, managers and others with the CAN/SDI Project.
2. Expansion of the user training program.
3. Utilization of new data bases to obtain wider coverage of scientific and technical literature.
4. Development of retrospective searching programs.

Simultaneously, and in keeping with the government's directive to develop in concert with existing information organizations a national STI system, there will be a continuing study of the project to determine:

1. The degree to which central processing of tapes can or should be expanded.
2. Those activities of the CAN/SDI Project which could be performed more efficiently by private agencies, e.g. tape searches and distribution of print-out.
3. The desirability of assigning to other agencies responsibility for providing within CAN/SDI, an SDI service covering subjects in which they have special competence.
4. The cost-effectiveness of the service.

Conclusion

This has been a rather hasty and incomplete review of one country's rather halting steps to provide a national SDI service through the adoption of an information handling technique which was developed as early as 1958. Any claims to originality and uniqueness rest primarily on the designing of a program which effectively overcomes the problem created by the incompatibility of data bases, and the utilization of both centralized and decentralized procedures to provide a current-awareness service to a relatively large population in one of the world's largest countries. Whatever success the program has achieved is due entirely to the ingenuity and efforts of the staffs of the NSL and of the NRC's Computation Centre, and to the cooperation and enthusiasm of those Canadian scientists who served as guinea pigs in the early stages of experimentation.

References

1. Science Council of Canada. A Policy for Scientific and Technical Information Dissemination (Report No. 6). Ottawa, Queen's Printer, 1969, 35 p.

2. Science Council of Canada. Scientific and Technical Information in Canada (Special Study No. 8). Ottawa, Queen's Printer, Part I, Part II (6 chapters) 1969.

3. National Science Library. Profile Design Manual: CAN/SDI Project. First Edition Revised. Ottawa, NRC, August 1969, 47 p.

4. Brown, Jack E. Canadian Engineering Communications Programs, National Engineering Information Conference Proc., Washington, June 24-25, 1969, p. 58-63.

5. Brown, Jack E. The CAN/SDI Project. "Special Libraries", 60: (No. 8), Oct. 1969, p. 501-9.

6. National Science Library of Canada. Annual Report and Guide to Resources and Services 1969-70. Ottawa, Information Canada, 1970, 39 p. (in English and French).

7. Brown, Jack E. The CAN/SDI Project: Training of Users. International Federation for Documentation, International Congress of Documentation - Proc., Buenos Aires, Sept. 21-24, 1970, 11 p.

8. Wolters, Peter H. and Brown, Jack E. CAN/SDI System: User Reaction to a Computerized Information System for Canadian Scientists and Technologists. "Canadian Library Journal", 28: (No. 1), Jan.-Feb. 1971, p. 20-3.

9. Heilik, J. Information Retrieval and MARC at the National Science Library. "Canadian Library Journal", 28: (No. 2), March-April 1971, p. 120-3.

10. Brown, Jack E. A National STI System for Canada. "National Research Council of Canada Newsletter", 3: (No. 1), Spring 1971, p. 3-4.

THE COSTS OF DOCUMENTATION SERVICES BASED ON MAGNETIC TAPES
Basis for the Choice of Price Policy

Scandinavian Conference in Copenhagen, 28th - 29th October 1970

Reported by: Helge J. Skov
Documentation Department
National Technological Library of Denmark

SYNOPSIS

The actual costs of running computer-based SDI services can be split up in 4 main components: Disbursement to the supplier of the tape system, working expenses of the documentation center, costs of the electronic data processing, and expenses connected with the procurement of the original material to the subscribers. The costs per profile for running 50 or 300 profiles in the Chemical Abstracts Condensates system are analysed, and distributed on the four components mentioned.

1. Background, Introduction

The Nordforsk committee of technical information, which is the expert group of Nordforsk in the field of Documentation and information, has for several years been concerned with the problem of cooperation between Denmark, Finland, Norway and Sweden in connection with computerized documentation services. As part of this cooperation a number of conferences and some smaller meetings have been held in order to discuss the problems of computer based documentation. Issues of current interest which have been discussed at these conferences, are the establishment of Scandinavian agreements regarding the procurement of documents, the use of tape services and related activities, the training of staff, and a common price policy. In this connection the need for a concrete numerical basis for the discussions has arisen.

The committee of technical information has therefore decided to arrange a conference, which through lectures, discussions and meetings of working groups should give a realistic picture of how much it actually costs to run a tape service. Furthermore the conference should serve the purpose of creating an analytical instrument for the policy-forming institutions of information and documentation in particular countries and form the basis for decisions related to the running of local or other tape services. The conference was planned and arranged by Mr. H.J. Skov, National Technological Library of Denmark.

2. The Programme, Participation

The programme and list of participants appears from appendix No. 1.

The conference was divided into three sessions: Approach to the problem, groupwork, and conclusions.

Altogether 22 persons took part in the conference, namely 5 from Denmark, 4 from Finland, 5 from Norway, 5 from Sweden and 3 from Nordforsk. Participation in the conference was based on an invitation of the Nordforsk committee of technical information after consultation, with the central institutions of scientific and technical information and documentation of each individual country.

3. Working Groups

As a concrete topic for the discussions of the conference a system as for instance Chemical Abstracts Condensates (CAC) with 50 or alternately 300 annual search profiles was chosen for elucidation from an economic point of view. It was assumed that the total costs could be divided into 4 main components: Disbursement to the supplier of the tape system, working expenses of the documentation center, costs of the electronic data processing, and expenses connected with the procurement of the original material to the subscribers.

After an introduction by Mr. Th. Franck and Mr. H.J. Skov short lectures of 20 minutes each were given by Mr. Anders Kallner, Mrs. Ch. Lindqvist, Mr. Peter Svare, and Mr. J. Brandrud as introduction to the group work.

4 working groups were established in accordance with the competence and wishes of each participant. Each of the 4 groups was given the task of elucidating one of the 4 mentioned cost components. At the end of the group work each group presented a written summary of its work. These summaries are included as appendices No. 2a-2d.

Plenary Meeting

The conference finished with a plenary meeting including a discussion of the results of the working groups and a conclusive debate under the leadership of Mr. Th. Franck. A summary of the conclusions appears in appendix No. 3.

According to the results of the conference and with the reservations stated in the appendix it is possible to put forward the following approximate distribution formula for the cost components of a system of the type CAC. The cost components are stated below in per cent per profile per year for 50 and 300 profiles (one profile comprises 20 search terms).

	S.Kr. (Swedish Crowns)		%	
Number of profiles:	50	300	50	300
I. Disbursement to the supplier.	600	150	15	4
II. Working expenses of the documentation center	1000	1000	25	30
III. Costs of the electronic data processing.	1400	1200	35	36
IV. Expenses connected with the procurement of original material.	1000	1000	25	30
Total:	4000	3350	100	100

The calculations, which were made in Swedish currency, revealed that the total amount for CAC per profile per year is between 3,700 and 4,400 Swedish Crowns for 50 profiles and between 3,000 and 3,800 Swedish Crowns for 300 profiles. The variations in the estimated costs of the documentation center and electronic data processing is not surprising in this very first attempt to make a cost estimate of components of documentation services. Later on comparative calculations may give less varied figures.

Approved 30th January 1971
Vibeke Ammundsen
President of the Committee of Technical Information, Nordforsk.

Appendix No. 1

Participants

Denmark:

Mrs. Inge Berg Hansen, M.Sc.
National Technological Library of Denmark
Øster Voldgade 10
DK-1350 Copenhagen K

Mr. Denis Dion, Chief adviser
I/S Datacentralen
Ved Stadsgraven 15
DK-2300 Copenhagen S

Mr. Holger Friis, Deputy Librarian
The University Library, Division II.
Nørre Alle 49
DK-2200 Copenhagen N

Mr. Helge Skov, M.Sc.
National Technological Library of Denmark
Documentation Department
Øster Voldgade 10
DK-1350 Copenhagen K

Mr. Peter Svare, M.Sc.
I/S Datacentralen
Ved Stadsgraven 15
DK-2300 Copenhagen S

Finland:

Mrs. Birgitta Holm, M.Sc.
Centrallaboratorium AB
Box 101 36
Helsingfors 10

Mr. Sauli Laitinen, M.Sc.
Helsinki Technological University Library
Otnäs

Mrs. Christina Lindqvist, M.Sc.
Finska Socker AB
Harju 1B 12 Kantvik

Mr. Ilkka Roman, M.Sc.
Helsinki Technological University Library
Otnäs

Norway:

Mr. J. Brandrud, Deputy Librarian
The National and University Library in Oslo
Drammensveien 42
Oslo 2

Mr. O. Brisner, Administrative secretary
Norwegian Industries Development Association
Forskningsveien 1
Oslo 3

Mr. H.M. Fagerli, Technical adviser
The National and University Library in Oslo
Drammensveien 42
Oslo 2

Mr. W. Holst, Manager
Norwegian Industries Development Association
Forskningsveien 1
Oslo 3

Mr. K. Thalberg, Library director
Library of the Technical University of Norway
Trondheim

Sweden: Mr. Hans Baude, Library director
The University Library in Linköping
Fack
581 83 Linköping 1

Mrs. M. Edström, Librarian
Royal Institute of Technology Library
100 44 Stockholm 70

Mr. A. Kallner, Dr., University lecturer
Royal Caroline Institute
Fack
104 01 Stockholm 60

Mr. Wolfram Uhlmann, M.Sc.
Försvarets Forskningsanstalt
FOA Index
104 50 Stockholm 80

Mr. Kjell Ölmestrand, Head of terminal
Inst. för ekonomi och statistik
Datasektionen
750 07 Uppsala 7

Nordforsk: Mr. Th. Franck, Managing director (Member of the board of directors of Nordforsk)
Danish Council for Scientific and Industrial Research
Hørsholm Kongevej 2
DK-2970 Hørsholm

VIII Commercially Available Services

Mrs. Vibeke Ammundsen, Library director (President of the Committee of Technical Information)
National Technological Library of Denmark
Øster Voldgade 10
DK-1350 Copenhagen K

Mr. Erling Hagen, M.Sc.
Nordforsk
Main secretariat of Nordforsk
Box 5103
102 43 Stockholm 5

Appendix 2a. Record from Working Group I:

Disbursement to the Supplier of the Tape System

Participants: Denis Dion
Helge J. Skov
Birgitta Holm
K. Thalberg
A. Kallner

Task

Construct and analyse a model of the disbursement to the supplier of the tape system taking into consideration the subscription and licence fee for e.g. references printed out (hit) both for current awareness services and for retrospective searches. Attempt also to estimate how much a supplier of a data base could be expected to demand as compensation for allowing storage of the data base on e.g. a disc to which many documentation centers or users might have direct access.

Account

In the following account we will confine ourselves to the conditions which affect Chemical Abstracts Service (CAS) and the data bases made by CAS, firstly Chemical Abstracts Condensates (CAC). The reason for this is the fact that CAS has a relatively long experience with tape systems in this connection and that their systems are relatively comprehensive. Furthermore the policy of CAS has been of guidance to many other documentation systems as for instance BIOSIS, COMPENDEX, ABIPC, etc.

CAS has asserted that in the future strict agreements should be made on the basis of a fixed base price and a certain royalty according to which CAS will receive an income in relation to the use of the system. The working group believes that the absolute level of the base price and the royalty cannot be discussed objectively in this connection, but is of the opinion that the present level is reasonable.

According to the working group the cost component of a profile with which the subscription contributes is as follows: Base price $ 4,400, freight, magnetic tapes etc., $ 1,000. A profile gives on average 26 references per search, which corresponds to a price per profile of approximately $ 120 and $ 30 at a load of 50 and 300 profiles, respectively. The above calculation does not take into consideration the possible use of the data base in connection with retrospective systems.

The group is also of the opinion that a leasing agreement with CAS as the formal owner of the tapes is acceptable.

The main objection to the proposed contracts concerns the restrictions according to the wishes of CAS regarding the use of the data bases. According to CAS every form of reproduction and issuing of the tape contents should be explicitly forbidden. CAS assumes that the customers of the documentation centers will accept this condition. As far as we can see, this will, to a great extent, prevent a rational exploitation of computerized documentation services without clearly reducing the possibilities for companies to sell their printed publications and magnetic tapes. We are of the opinion that these two documentation systems supplement each other and will not compete when within a few years a less dramatical situation can be foreseen.

Consequently the working group suggests that a future discussion concerning the formulation of the contract should aim at the abolition of these limitations. We suggest that three main lines of action be tested; these are listed below without any relevant priority:

1. Another form of financial agreement should be pursued. It is obvious that the Scandinavian countries in common or separately could take part in delivering input to the system. This could be done for instance by the Scandinavian countries undertaking (like U.K.) the editing and drawing up of abstracts and indexing of articles which are published in the Scandinavian technical journals.

2. Negotiations should be started with the purpose of establishing a price at which the producer (CAS) is willing to sell all material without the present limitations of use. In this connection several specific partial solutions could be considered, which allow e.g. the building up of retrospective files, direct access, the printing out of text on the tapes, etc.

3. An adequate indexing of scientific literature is not only of international interest, but also to a high degree an international responsibility. This should not be supervised by one single national institute or organization, instead one should aim at attaining a "multinational indexing-institute" where different nations for instance the Scandinavian countries, in common or separately, commit themselves to grant a certain financial support to the indexing institute without a direct relationship to the use of the institute. An estimation of this financial support could serve as a feature in the discussion of item 2 above.

28.10.1970
A. Kallner

Appendix 2b. Record from Working Group II:
Working Expenses of the Documentation Center

Participants: Holger Friis
Sauli Laitinen
Christina Lindqvist
W. Holst
Malin Edström

Task

Construct and analyse a model of the working expenses of the documentation center. Which and how big a staff is necessary and how much will it cost? Appraise the office rent, travel expenses, instruction and teaching work in connection with a modern and progressive center. Estimate also the costs involved in the current training of the staff required.

Account

Since none of the members of the group have experiences from Chemical Abstracts we had to base our conclusions on the experience of KTH (Royal Institute of Technology Library) in Stockholm with its 12 tape systems and THB (Helsinki Technological University Library) with PANDEX as well as Danish experience with MEDLARS.

Regarding the alternatives, 50 and 300 profiles, the group found that a total of 50 profiles would not be sufficient to form the basis for a documentation service based on tapes except for an introductory period.

The group has summed up the staff requirements as follows:

For the profile formulation: 2 - 4 hours/profile.

In order to raise the number of profiles to 300 when starting from scratch 1/2 manyear is required.

In order to maintain these profiles in connection with cancellations, new subscriptions and changes, it is necessary to have:

minimum: 1 information officer
maximum: 2 information officers

with the supposition that 50% of the time is used for work with the profiles while the other 50% is spent in keeping contact with clients including time for training.

In order to follow up this work by punching profiles, copying and sending out printouts each week, as well as invoicing and procuring articles for copying it is necessary to employ:

minimum: 1 secretary
maximum: 2 secretaries

VIII Commercially Available Services

The Necessary sales work sums up to 1 manyear, however, the group has included this work in an overhead of 100%.

Once-for-all procurement of brochures and leaflets	D. kr.	10.000,00
Annual demand for brochures	D. kr.	5.000,00

The salary level is clearly different in the four Scandinavian countries. The group has agreed upon the following salary base (D.kr. = Danish Crowns):

Documentalist (3 years experience)	70.000,00
Secretary	35.000,00
Sales staff	70.000,00

The staff costs amount to (D.kr.):

The working up of 300 profiles	1/2 manyear	35.000,00
Maintenance of 300 profiles	minimum 1 -	70.000,00
	maximum 2 -	140.000,00
Assistance	minimum 1 secretary	35.000,00
	maximum 2 secretaries	70.000,00
Instructive material, brochures, etc.		5.000,00

Total:	Minimum (D.kr.)	Maximum (D.kr.)
The building up	35.000,00	35.000,00
Maintenance	70.000,00	140.000,00
Secretary	35.000,00	70.000,00
	140.000,00	245.000,00
Overhead 100%	140.000,00	245.000,00
	280.000,00	490.000,00
Brochures	5.000,00	5.000,00
Travelling	10.000,00	10.000,00
	295.000,00	505.000,00

With 300 profiles this gives a cost per profile of:

minimum: D.kr. 1,000 (S.kr. 750 with a rate of exchange of 0,75)

maximum: D.kr. 1,700 (S.kr. 1,250 with a rate of exchange of 0,75)

28.10.1970
W. Holst

Appendix 2c. Record from Working Group III: Costs of the Electronic Data Processing

Participants: Peter Svare
Ilkka Roman
O. Brisner
H.M. Fagerli
W. Uhlmann
K. Öhlmestrand

Task

Construct and analyse a model of the costs related to the operation of hardware and the development of software as for instance search programs. Furthermore evaluate from an economic point of view whether print-outs of references with long abstracts are more profitable than print-outs which only contain author, title, keywords, and journal reference.

Make also an appraisal of the advantages and drawbacks of having a computer center of its own in comparison to running the tapes on a bigger center on time sharing basis.

Account

Costs of development

Reformating programs:

A general information retrieval system requires one reformating program for each tape type. The purpose of the reformating program is to edit data from the original tape to the format of the search program.

Planning, programming and testing of a reformating program costs about 5.000 - 10.000 S.kr. (Swedish Crowns).

Other information systems:

The other information systems consist of programs for reading in, searching, sorting and printing out. The costs for the development of such a system are highly dependent on how complicated a system is required. As an indication of the costs we have chosen the figures corresponding to CORSAIR and TELETEXT.

The development of CORSAIR has to date cost 1/2 – 3/4 million Swedish Crowns.

The development of TELETEXT has to date cost approximately 1/2 million Swedish Crowns.

The present annual development costs are for CORSAIR 100-200,000 S.kr. and for TELETEXT/BATCH approximately 40,000 S.kr.

The Working Expenses

The working expenses cover the use of hardware and operator assistance during the runs. However, the preparation of search runs is not included since it is assumed that the preparation is done at the documentation center.

Costs connected with the reformating:

The magnitude is 100 S.kr. for 1,000 references of 1,000 characters each.

Costs of searching and printing out:

The scanning of a tape with 5,000 references and printing out of 100 lines per profile costs about 20 - 25 S.kr. per profile.

The following formula can be used for the calculation of the cost magnitude in connection with other numbers of search terms:

$$\text{Costs: } (120 + 0{,}03 \cdot T + 0{,}00013 \cdot S \cdot T + 0{,}1 \cdot L) \cdot 0{,}83 \text{ S.kr.}$$

T = number of references searched
S = number of search terms
L = number of lines printed out.

Compared with the other costs of the data processing the number of lines printed out is of less importance and should therefore be determined from a documentation point of view.

Regarding the question whether a documentation center should have its own computer or not, one can state that generally there are considerable advantages connected with the running of big computers.

28.10.1970
Peter Svare

Note:

The total cost is:

for 50 profiles (CAC): 1,300 - 1,550 S.kr. yearly per profile.
for 300 profiles (CAC): 1,100 - 1,350 S.kr. yearly per profile.

E. Hagen

Appendix 2d. Record from Working Group IV:

Expenses Connected with the Procurement of the Original Material to the Subscribers

Participants: Vibeke Ammundsen
Inge Berg Hansen
J. Brandrud
H. Baude

Task

Construct and analyse a model of the costs connected with the procurement of copies of journal articles, patents, etc., often from libraries far away. In this connection one should also take into account the consequences of more rigorous copyright conditions according to the trends in the big countries, since efficient SDI-services will reduce the industrial users' requirements for having their own subscriptions to periodicals.

Account

Fundamental considerations

1. Free copying for subscribers is considered unrealistic.

2. Payment is required only of the actual costs excluding labour costs.

3. The documentation center should have a reasonable coverage in its own collection of literature in relation to the service it offers. For reasons of time and economy one should aim at Scandinavian coverage of the relevant journal literature to the greatest extent possible; this implies the existence in one form or another of a certain survey of the stock of periodicals at the scientific libraries in Scandinavia.

   The question of the translation services should be evaluated separately.

4. Each particular Scandinavian country should as far as possible exploit its own resources, but in such a way that specialized libraries should still be able to establish direct contact across the borders.

Cost calculations

The expenses connected with the procurement of original material is difficult to calculate exactly and will vary from subject to subject. The expenses are also dependent on how the material is procured. Time consumed on the ordering side could on an average be estimated at 1/2 hour per order. The calculations are based on the technical or scientific articles having an average length of 8 pages according to the National Technological Library of Denmark. Furthermore (for practical reasons) the calculations are based on the prices of National Lending Library. The expenses connected with the procurement of a copy of an article can then be characterized as follows:

Labour costs: 1/2 hour at 30 N.kr.	15 N.kr.	(Norwegian Crowns)
Payment for copies	5 N.kr.	
Various expenses (postage etc.)	2 N.kr	
Minimum cost per copy of an article:	22 N.kr.	

It is also possible to calculate the costs per profile. We lack accessible data regarding Chemical Abstracts Condensates, but the experience of the National Technological Library of Denmark from their work with Chemical Titles shows that on average 180 copies (pages) per profile (1 profile = 20 search terms) per year are ordered. This corresponds to approx. 22 articles of 8 pages each. Accordingly a profile will cost approximately 500 N.kr. per year. The reservation is made that many of the users have access to big collections of their own. If <u>all</u> orders for copies were handled by the documentation center one would have to allow for an addition of 25%, which would give a price per profile of approximately 625 N.kr. per year.

If the calculations were transferred to CAC then at least a doubling, that is at least 1,250 N.kr. per profile, can be expected.

If one also takes into account a possible future <u>copyright fee</u> of e.g. 15% (of the copy costs), the price per profile for CT amounts to approximately 660 N.kr. per year.

On the assumption that the expenses of labour should be covered by the community the customer would pay 140 N.kr. of an annual cost of 625 N.kr. per profile, while the documentation center pays the rest (485 N.kr).

In the above calculations it has not been found relevant to include the costs of the stock of periodicals.

<u>Copyright problems</u>

More rigorous copyright conditions would have a restrictive influence on the utilization of research results. It is desirable to reach decisions which do not stimulate violations of the rules by the restrictions they impose. The establishment of an obligation to pay the publisher/author <u>a fee for copying</u> is recommended by analogy with the arrangement existing for musical works. This would give the periodicals a compensation for sales losses and probably also prevent misuse.

The group has not found it possible to take up a position as regards serious copyright restrictions hindering or delaying copying for research purposes.

28.10.1970
J. Brandrud

Appendix 3

Summary of the Records from the Working Groups. Conclusion

The results of the working groups are stated below as costs in S.kr. (Swedish Crowns) per profile per year for a tape system like CAC with a load of 50 alternatively 300 profiles (one profile comprises 20 search terms).

The comments from the plenary meeting and the total costs are mentioned below.

		CAC costs per profile per year	
	Number of profiles:	50	300
I.	Disbursement to the supplier.	600	150
II.	Working expenses of the documentation center.	750 - 1250	750 - 1250
III.	Costs of the electronic data processing.	1300 - 1550	1100 - 1350
lV.	Expenses connected with the procurement of original material.	1000	1000
	Total:	3650 - 4400	3000 - 3750

Re. I:

The utilization of the database in connection with a retrospective system has not been taken into consideration in the calculations.

Re. II:

The calculations have only been carried out for 300 profiles, since 50 profiles is considered too small a quantity to constitute a proper economic foundation for a tape-based service of one's own. In order to obtain a realistic standard of reference the cost interval corresponding to 300 profiles is also used in the case of the 50 profiles.

Re. III:

The calculations are presented as costs evaluated in the case of starting a new establishment. However, if leased computer time is to be used on a commercial basis then the prices will vary within relatively wide limits; accordingly a certain margin will probably be applicable to this cost component.

On the basis of accounts for the period 1968-70 an average annual cost per profile for CAC of 803 S.kr., corresponding to a load of 276 profiles (of 25 search terms each) has been indicated by the Swedish participants. The market price for computer time was in this case 755 S.kr. per hour.

Re. IV:

In this case the costs per profile are dependent on the number of profiles. The numbers given are based on Chemical Titles, while CAC is expected to imply a doubling of the copies ordered.

The total costs are here rounded off to 1000 S.kr. per profile per year.

Re. Total:

The stated sums of the components should not be taken as being based on accurate and fully valid calculations but rather as an expression of a magnitude. By stating approximate average values for each cost component one should therefore be able to simplify the account without introducing new errors of importance. If this is done the cost distribution in S.kr. and % becomes:

		CAC costs per profile per year			
		S.kr.		%	
	Number of profiles:	50	300	50	300
I.	Disbursement to the supplier	600	150	15	4
II.	Working expenses of the documentation center.	1000	1000	25	30
III.	Costs of the electronic data processing.	1400	1200	35	36
IV.	Expenses connected with the procurement of original material.	1000	1000	25	30
	Total:	4000	3350	100	100

The distribution in % which mainly becomes the same even though the maximum/minimum values for II and III are used shows that the disbursement to the supplier of the tape system constitutes only a small part of the total costs for a load of 300 profiles. The computer costs for the number of profiles included in these calculations amount to approximately 1/3 of the costs. The working expenses of the documentation center and the expences connected with the procurement of the original material are according to the above calculations, approximately the same, and constitute an increasing part of the total costs for an increasing number of profiles.

Without drawing too far-reaching conclusions from this distribution it is reasonable to assume that the conditions with regard to other commercially available documentation services based on magnetic tapes do not deviate essentially from the service described above.

Stockholm, November 9, 1970.
Erling Hagen, M.Sc.

COMMERCIALLY AVAILABLE INFORMATION SERVICES

B. Doudnikoff
President
Dataflow Systems Inc.
Bethesda, Maryland, USA

SYNOPSIS

To start a comprehensive new library or information center usually is an expensive operation. In fact, the expansion of coverage in an existing library or information center is also linked with relative high cost. A number of "packages" are available to help in these areas. These "packages" come in various forms, including magnetic tapes for processing on a computer, microform, hard-copy (paper or forms) and combinations of these.

Introduction

In an effort to assist librarians and information specialists (and to make some money) numerous organizations have assembled "packages." Presented here is a selected sample of these packages. Unfortunately, some very good ones are excluded. However, if the concept behind these "packages" is of interest to the reader it is relatively simple to pursue specifics in a particular discipline area.

To present these sample "packages" the medium versus the discipline was considered. The medium won out. Therefore, if discipline is of key interest, selections must be made from each medium. Microform, computer magnetic tapes, and hard-copy publications are the media. For the purpose of this description the hard-copy services are being de-emphasized.

Microforms

The miniaturization of information and data onto a variety of microforms has been done for many years on a custom basis. As organizations realized the cost economies of mass production of microforms a new era of micropublishing has evolved.

The specific microform (16mm reel or cartridge/magazine, microfiche, etc.) will not be discussed here. If detailed information on the media and the required reader and reader/printer is needed, it can be obtained from the firms that sell the hardware or from the firms that sell the software (the "packages").

Listings of the selected microform packages are shown as follows:

BIOLOGICAL ABSTRACTS ON MICROFILM

BioSciences Information Service of Biological Abstracts, Philadelphia, Pennsylvania. 16mm cartridge, bar code, image count, serial sequence, external book indexes (computer produced). Commercial lease. 125 cartridges. 1926 to present. Updated bimonthly.

Two million abstracts of articles and reports in life science research, $1600 first year with back files, $600 thereafter for current films, but the subscriber must also purchase the current hard-copy index for $800 per year.

Biology	Biostatistics	Weapon Effects
Biochemistry	Biophysics	Bioengineering
Behavioral Biology	Ecology	Botany
Zoology	Forestry	Food Technology

Order from: BioSciences Information Service of Biological Abstracts, 2100 Arch Street, Philadelphia, Pennsylvania 19103, Attn: Dr. John H. Mason. Telephone (215) LO9-1100 extension 37 or 38.

"CARDSET" LIBRARY OF CONGRESS CATALOGING DATA ON MICROFILM

Information Design, Inc., Menlo Park, California. 16mm cartridge. Separate index, also on 16mm microfilm. Commercial subscription. Over 100 cartridges. Current, updated every other week.

Covers all English Language cataloging included in MARC (see magnetic tape description). Index stations can be leased separately. $3850 per year for total system.

Additional information from: Information Design, Inc., Menlo Park, California 94025, Attn: Mr. Brett Butler. Telephone (415) 369-2962.

CHEMICAL ABSTRACTS ON MICROFILM

Chemical Abstracts Service, Ohio State University, Columbus, Ohio. 16mm cartridges, bar code, accession number sequence, external book index. Commercial lease. Current, updated 8-10 times per year.

An English language abstracting and indexing service, Chemical Abstracts covers chemistry and chemical engineering wordwide. Estimated 1970 volume is 270,000 abstracts from all sources of chemical literature. Indexes not microfilmed. Chemical Abstracts Service is a division of the American Chemical Society.

Chemical Abstracts
Chemistry
Chemical Engineering

Prices effective in 1971, $2,200 for next twelve months including back issues from 1907 to present, $1,550 for succeeding years, to current subscribers of printed issues of Chemical Abstracts.

Additional information from: Chemical Abstracts Service, Ohio State University, Columbus, Ohio 43210. Attn: Marketing Department. Telephone (614) 293-1929.

CUMULATIVE MICROFILM INDEX TO USGRDR

Clearinghouse for Federal Scientific and Technical Information, U.S. Department of Commerce, Springfield, Virginia. 16mm reels or cartridges, sequential order. Government sale. 20 reels. Reissued in 1969.

The Clearinghouse has been experimenting with an index on 16mm microfilm which cumulates the United States Government Research and Development Reports Indexes from July 1964 to July 1969. This index was developed on an experimental basis to use in handling general reference searches received by the Clearinghouse.

The result is a product containing over 150,000 titles that represents the first comprehensive computer-generated index to the U.S. Government report literature. It allows reference searchers to browse this literature quickly for retrieval purposes.

While the file is not polished, its continuity of coverage may make it useful to librarians and technical information specialists working in the information field. This file is now available to the general public.

The Index is in two segments - July 1964 - December 1967 (Reels 1-13) and January 1968 - June 1969 (Reels 14-20).

Technical Reports	Chemistry	Chemical Engineering
Electronics & Electrical Engineering	Mechanical Engineering	Aeronautics
Social Sciences	Mathematical Sciences	Astrophysics
Space Technology		Physics

The Microfilm Index is available in the following format:

Twenty 100-foot reels at $100
Twenty 100-foot 3M cartridges at $135
Twenty 100-foot Recordak cartridges at $135

Order from: Clearinghouse for Federal Scientific and Technical Information, U.S. Department of Commerce, Springfield, Virginia 22151. Inquiries: Telephone (703) 321-8560.

IDEP (INTERAGENCY DATA EXCHANGE PROGRAM)

Army/NASA IDEP Office, U.S. Army Missile Command, Redstone Arsenal, Alabama. 16mm reel or cartridge, bar code, accession number sequence, optical coincidence and computer listing type index. Limited availability to qualified agencies and contractors. About 285 cartridges. Updated quarterly.

Covers aerospace and missile test reports and specifications of components and parts as prepared by participants other than original manufacturers. Also covers test procedures, reliability history and failure analysis. Contains about 30,000 test reports and 15,000 calibration procedures. Uses optical coincidence index, and book index by subject. Sponsored by Army, Navy, Air Force, and NASA. Other IDEP Offices: Fleet Missile Systems Analysis and Evaluation Group (Code E-63), Corona, California 91720, and Los Angeles Air Force Station, Code HQ SAMSO (SMSDI), Los Angeles, California 90045. Limited to agencies and contractors in aerospace and weapon systems area participating in actual testing.

Components	Specifications	Reliability
Missiles	Aircraft	Tests
Performance Tests	Mechanical Parts	Failure
Mechanical Properties	Calibration Procedures	Test Equipment
Measuring Instruments		

For additional information about IDEP, contact: U.S. Army Missile Command, Army/NASA IDEP Office, Redstone Scientific Information Center, AMSMI-RBP, Redstone Arsenal, Huntsville, Alabama 35809. Telephone: Area Code (205) 876-8720

"INFACT" COLLEGE CATALOGS ON MICROFILM

Dataflow Systems, Inc., Bethesda, Maryland. 35mm roll film. External "peek-a-boo" index. Commercial subscription. 20 rolls. Current, updated quarterly.

Covers about 500 leading U.S. colleges and universities undergraduate catalogs. Index, "College Suggestor" also sold separately. Price range from $75 per year (regional coverage) to $495 per year (total selected coverage).

Additional information from: Dataflow Systems,Inc. 7758 Wisconsin Avenue, Bethesda, Maryland 20014. Attn: J.B. Malcom. Telephone (301) 654-9133.

SHOWCASE MICROFILM LIBRARY FOR CONSTRUCTION (SML-C)

Showcase Corporation, Detroit, Michigan. 16mm cartridge, sequential frame numbers, book and filmed index. Commercial lease. 100 cartridges. Current, updated quarterly.

A complete microfilm library designed specifically for the construction industry and design profession, containing manufacturers' catalogs, selected specifications from the U.S. Government (including pertinent Federal Specifications, U.S. Department of Commerce Specifications, Federal Housing Specifications and Standards, Army Corps of Engineers Guide Specifications, and others), and relevant information from industry associations, institutes, and government agencies. The file is indexed by product, keyword, trade name, and manufacturer's name.

Civil Engineering	Vendor Catalogs	Specifications
Construction Equipment	Construction Materials	Construction

Order from: Showcase Corporation, Showcase Building, 6230 John R. Street, Detroit, Michigan 48202, on annual lease for $2,800.

UNITERM INDEX CHEMICAL PATENTS ON MICROFILM

IFI/Plenum Data Corporation, Washington, D.C. 35mm and 16mm reel accession number sequence. External dual dictionary index and magnetic tape search system. Commercial sale. 1959 to current. Updated monthly.

All U.S. Chemical patents from Official Gazette classification and all chemically related patents from other classes. Filmed in IFI accession number sequence by years. Available in the following series to complete patents contained in the Uniterm Index:

1959 - 1963	(58,113 patents)	35mm reels	$1,850.
1964 - 1966	(50,007 patents)	16mm reels	900.
1967	(18,964 patents)	16mm reels	515.
1968	(18,000 patents)	16mm reels	515.
1969	(20,000 patents)	16mm reels	515.
1970	(20,000 patents)	16mm reels	515.

Also available:

Patent Claims contained in the Uniterm Index:

1950 - 1969	reel or cartridge	16mm	$1,350.
1970	reel or cartridge	16mm	200.

Minor Terms contained in the Uniterm Index:

1950 - 1963 (282,000 compounds)
6 cartridges 16mm $2,200.

Chemistry
Chemical Engineering
Patents

Order from: IFI/Plenum Data Corporation, 1000 Connecticut Avenue, N.W., Washington, D.C. 20036. Attn: H. Alcock Telephone: (202) 296-4936

U.S. PATENT SUBSCRIPTION SERVICE

Clearinghouse for Federal Scientific and Technical Information, Code 52.12 (Subscription Sales), U.S. Department of Commerce, Springfield, Virginia. 16mm reels, bar code, patent number sequence, no index provided. Annual subscription. About 250 reels per year. 1966 to present. Updated as required.

All U.S. patents are contained in the Offical Gazette, which serves as an index. Reels issued as accumulated, about one month after announcement. Back years from 1966 are available at $895 per year. Annual subscription all classes, is $895 per year. General and Mechanical category only is $600 per year. Electrical category is $400 per year and Chemical category only is $300 per year.

Patents — Mechanical Engineering
Chemistry — Electronics & Electrical Engineering
Chemical Engineering

Order from: Clearinghouse for Federal Scientific and Technical Information, Code 52.12, U.S. Department of Commerce, Springfield, Virginia 22151. Attn: Mr. Foster. Telephone:(703) 321-8507

VIII Commercially Available Services

VSMF DESIGN ENGINEERING FILE

Information Handling Services, Division of Indian Head, Inc., Englewood, Colorado. 16mm cartridge, bar code number frames, book and film index. Commercial lease. About 220 cartridges. Current, reissued quarterly.

Catalog data, specifications, drawings, and test reports on parts, materials and services used in industry. Indexed and filmed by product groupings.

Vendor Catalogs	Specifications	Engineering Drawings
Vendor Part Numbers	Test Reports	Mechanical Engineering
Electronics & Electrical Engineering	Components	Mechanical Parts

Available on monthly lease for $480 per month, or annually for $5360.
Order from: Information Handling Services, 5500 South Valentia Way, Englewood, Colorado 80110.

VSMF MILITARY SPECIFICATION FILE

Information Handling Service, Division of Indian Head, Inc., Englewood, Colorado. 16mm cartridge, bar code numbered frames, filmed index. Commercial lease. About 205 cartridges. Current, updated monthly.

Specifications and book standards used by industry, covering 483 Federal Supply Classifications, MIL SPECS, QPL's, MIL STDS, JANS, FED SPECS, and FED STDS, in six military specification files. Numeric index on film. Also available in 8mm. Historical file also available.

Specifications	Federal Specifications	Military Specifications
Military Standards	Federal Standards	Standards
Qualified Product Lists		

Available on 16mm on annual lease as follows:

Assemblies	- $450
Electrical	- 1120
Instruments	- 450
Mechanical	- 1065
Procedures	- 165
All	- 3350

Order from: Information Handling Services, 5500 South Valentia Way, Englewood, Colorado 80110.

VSMF MILITARY STANDARDS FILE

Information Handling Services, Division of Indian Head, Inc., Englewood, Colorado. 16mm cartridge, bar code, numbered frames, filmed numeric index. Commercial lease. 10 cartridges. Current, updated bimonthly.

Microfilm cartridges of MS, AN, AND, NASA Standard Parts; MIL-D-1000, MIL Handbook H4-1, H4-2 and 300 on ground support equipment.

Standards
Military Standards

Available on 16mm on an annual lease for $560.

Order from: Information Handling Services, 5500 South Valentia, Englewood, Colorado 80110.

MAGNETIC TAPES (FOR COMPUTERS)

Instead of disseminating information in microform some organizations have chosen the media of magnetic tape. Unlike, microform, magnetic tape information/ data must be processed on a computer to make it useable/readable. However, this added complexity also is the key asset of magnetic tape services. The information is "tailored" to local needs on the local computer. The microfilm is non-flexible, but is directly readable with minimal need for machines (outside of the relatively simple microform readers and reader/printer).

A quick review of the better known services is as follows:

BA PREVIEWS

BioSciences Information Service of Biological Abstracts, 2100 Arch Street, Philadelphia, Pennsylvania.

Subject Matter: Research literature of the world dealing with life sciences in any of their manifestations.

Types of Source Items Input: Approximately 8000 journals are reviewed for input, and an average of 4500 of these are productive during a calendar year. Of these journals 69% of them are English language publications. Journals represent 84% of the total input. Monographs, published proceedings, theses and selected abstracts from the Referativnyi Zhurnal Biologia account for the remainder.

Subject Analysis/Indexing: Items are assigned an average of 19 index terms or descriptors. Words from the title and abstract are used as keywords or descriptors. In addition, the title is enriched by adding descriptors and uncontrolled descriptors are selected from the text. No thesarus is used in indexing.

Searchable Data Elements: Keywords, CROSS index (subject headings), Biosystematic index, author, Journal source (CODEN).

Time Span Available: 2 years in BA Previews format. (11 years in machine readable form for inhouse searches).

Frequency of Tape Issue: 3 tapes per month (BA semimonthly, BioI monthly)

Average Number of Source Items Cited Per Tape: 6400

Subscription Cost or Leasing Details:
Tapes are available on lease. $3800 (includes 2 monthly tapes from BA of 5800 citations each and 1 tape BioI of 7500 citations) lease price on site. Royalty for additional customer use is negotiable. Those interested should contact the

source.

Software Availability: None

Type of Inhouse Service Offered:
CLASS (Current Literature Alerting Search Service), an SDI-type service based on each tape as it becomes available. The cost for this service is $100/year per search profile. Retrospective Searches are available. The cost for this service is $150 per search (with the right reserved to refuse a search at this price if it presents extraordinary problems).

Publications Produced from Base by Originator:
B.A.S.I.C., BioResearch Index, Annual Cumulative Indexes to Biological Abstracts, BioResearch Index, Abstracts of Mycology, Abstracts of Entomology.

Additional Information from: BioSciences Information Service of Biological Abstracts, 2100 Arch Street, Philadelphia, Pennsylvania, 19103, Attn: Mrs. Phyllis V. Parkins. Telephone (215) 569-1100.

PANDEX - CURRENT INDEX TO SCIENTIFIC & TECHNICAL LITERATURE

CCM Information Corporation, New York, New York

Subject Matter: 2100 scientific, technical and medical journals. 6000 scientific technical books. 50,000 U.S. Government technical reports.

Types of Source Items Input: 2100 journals have all articles contained in them entered into the data base. Of these journals 70% are English language publications; 30% non-English. Journal articles account for 80% of the data base; 18% of the input is taken from government reports; 2% are monographs.

Subject Analysis/Indexing: Each item has an average of 6 terms assigned to it. Terms are selected from a controlled thesarus. In addition to the indexing, titles entered into the record are "enriched" by adding descriptive words to the published title.

Time Span Available: 1966

Frequency of Tape Issue: Weekly

Average Number of Source Items Cited Per Tape: 5000

Subscription Cost or Leasing Details:
Educational, industrial, government or quasi-government organizations using the tape for inhouse use: $6500. The same price applies to non-profit regional information centers providing profile searches to customers at stated rates. Multi-user groups providing SDI service from central computer group to a group of companies in industry or a group of universities should contact CCM for price details. Profit information centers providing profile searches should also contact CCM.

Software Availability: COBOL programs print-out and SDI. IBM 360/DOS. 32K core. 7 or 9 track 800 BPI. Coding is EBCDIC, BCDIC, or ASCII. Fixed field or

MARC II format.

Publications Provided from Base by Originator:
Pandex -- Current Index to Scientific & Technical Literature. Bi-weekly publication. Free hard copy is supplied to tape subscribers.

Additional Information from: CCM Information Corporation, 909 Third Avenue, New York, New York 10022, Attn: Richard P. Kollin. Telephone (212) 935-3087.

CA-CONDENSATES

Chemical Abstracts Service, The Ohio State University, Columbus, Ohio.

Subject Matter: Bibliographic citations plus Keyword Index phrases from Chemical Abstracts issues.

Types of Source Items Input: The tape is to provide searchable data items from the corresponding issue of Chemical Abstracts. There is an accompanying printed version of the corresponding Chemical Abstracts issue.

Subject Analysis/Indexing: Titles of articles and patents are enriched by adding descriptors. Words from the title and abstracts are used as descriptors. Uncontrolled descriptors or keywords from text are selected and added to the record.

Searchable Data Elements: Words in title, keyword phrases, author name(s), and journal CODEN.

Abstracts: None. However, the record does include shorter phrases which may be used in lieu of abstracts.

Time Span Available: Available from July, 1968.

Frequency of Tape Issue: Weekly.

Average Number of Source Items Cited per Tape: 5000 (250,000 new articles and patents each year).

Subscription Cost or Leasing Details:
Tapes are available on lease-subscription at $4400/year plus costs of tapes and mailing. Interested subscribers should contact CAS as to type of lease/license for which they qualify.

Software Availability: Contact the source regarding this item.

Type of Inhouse Service Offered: Search of file based on customers search profile. Cost: $4400 plus charge for computer time used.

Note: Further details describing services are included in Information Services 1971. Available upon request from Chemical Abstract Services.

Additional Information from: Chemical Abstracts Service, The Ohio State University, Columbus, Ohio 43210, Attn: Marketing Department, Telephone

(614) 422-1929.

U.S. GOVERNMENT RESEARCH & DEVELOPMENT REPORTS

National Technical Information Service, Springfield, Virginia

Subject Matter: Broad coverage of all sciences, technical subjects and social sciences.

Types of Source Items Input: Many of the documents announced in USGRDR are received from the Department of Defense (DOD), the National Aeronautics and Space Administration (NASA) and the Atomic Energy Commission (AEC). Contributions from other federal departments make up the balance of input.

Of the total input, approximately 16% is reprints resulting from federally sponsored research and published in professional journals; 10% is translations; and 5% is foreign language documents.

Subject Analysis/Indexing: An average of 10 index terms selected from a thesarus of controlled descriptors, are assigned to each document record. Uncontrolled keywords from the text are also part of the unit record.

Searchable Data Elements: See page of data elements which follows:

Abstracts: Abstracts provided for 73% of items.

Time Span Available: Tapes available beginning with January 1970.

Frequency of Tape Issue: Twice monthly.

Average Number of Source Items Cited per Tape: 1800.

Subscription Cost or Leasing Details:
$1500/year for domestic subscription.
$1835 for foreign subscriptions, including air mail costs.
Tapes are available only on subscription basis.
Test tape of USGRDR v. 69, no. 13, available at $25/tape.

Software Availability: No programs for using the data will be made available by the Service.

Type of Inhouse Service Offered: Limited search service now offered. Subject search done on only 20% of subject terms considered most central to document. Title listing and accession number provided.

Publications Produced from Base by Originator:
USGRDR, Clearinghouse Announcements in Science & Technology, U.S. Government Research & Development Reports Index (USGRDR-I).

Further data in CFSTI brochure of 20 March 1970 which includes: Clearinghouse Announcement Journal Available on Magnetic Tape (Flyers); Subjects Covered; Data Elements; Description and Format for Clearinghouse USGRDR Magnetic Tape; Order

Form for Sample Tape.

. Additional Information from: National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22151, Attn: Marvin E. Wilson.

COMPENDEX (Computerized Engineering Index)

Engineering Index, Inc., New York, New York

. Subject Matter: All fields of engineering and certain fields of applied science and management, pertinent quality research applications literature; assembles the various engineering disciplines side by side with interconnecting cross reference.

Types of Source Input: More than 3500 sources of engineering literature -- professional and trade journals, publications of engineering organizations, papers from conferences and symposia, books and other documents.

Subject Analysis/Indexing: Items are assigned main headings, subheadings, and cross references, as required, from "SHE" (Subject Headings for Engineering), an authority list in use by Ei editorial staff. An average of 6 access terms (including subject headings and subheadings) are assigned input items. A maximum of 5 free language terms may be assigned.

Searchable Data Elements: Users have the option of searching the complete record (full text search) or searching selectively on chosen data elements; Subject heading and subheading, document ID number, author(s), Ei abstract number, codes assigned to the CARD-A-LERT service divisions, on most records the access words (which are cross references chosen for the printed version of Compendex), and on most records the free language terms.

Abstracts: Abstracts are included.

Time Span Available: 1969-1970 and subsequent years.

Frequency of Tape Issue: Monthly.

Average Number of Source Items Cited per Tape: 6000.

Subscription Cost or Leasing Details:
Tapes are available on lease only with subscription to Engineering Index (monthly and annual issue). Cost for combined package $6800/year (for 1970: $6000 for lease; $500 for combined annual and monthly indexes; $300 for 12 tape reels). Commercial (for profit) organizations should contact Ei to determine Compendex) availability and limitations on its use.

Software Availability: None. (Tape is in TEXT-PAC input format. TEXT-PAC is an IBM program for text search; available from IBM).

Publications Produced from Base by Originator:
None.

For Additional Information: Compendex brochure, Engineering Index, Inc., 345 East 47th Street, New York, New York 10017, Attn: John W. Carrigy, Telephone

(212) 752-6800. The brochure contains input format, data elements information on TEXT-PAC and various details regarding this service.

COMBINED SOURCE AND CITATION DATA TAPE

Institute for Scientific Information, Philadelphia, Pennsylvania

Subject Matter: Broad interdisciplinary coverage of journal literature, including the primary journals of basic and applied science, engineering and technology, medicine, psychology and psychiatry, and the behavioral sciences. (Data appears on tape within 15 days after publication.)

Types of Source Items Input: 2180 journals are input cover to cover and these articles comprise 98% of the data base. The remaining 2% of the tape is from monographs, published proceedings, theses, etc. Certain regularly apearing symposia reports of proceedings, are routinely processed as they are published. 50% of the base is from English language publications and 50% from non-English language publications.

Subjects Analysis/Indexing: "Indexing" in the sense of assigning either keywords or descriptors is not done. It is suggested that in lieu of this field the reader consider searchable data elements.

Searchable Data Elements: It should be noted that this tape includes not only the title of the original paper but all references cited in that paper. Searchable elements therefore include: 1) Citation line identifying any earlier published paper, book, patent, etc. 2) Words in a title including initial floating and terminal words stems or phrases. 3) Author currently publishing (the source author). 4) Authors whose earlier works have been cited (all cited authors). 5) The current journal title, i.e. the journal in which the original article appeared. 6) Earlier issue(s) of any journal(s) which has been cited in the references of this particular paper (cited journal). 7) Address where the work was done, i.e., the addresses of all authors.

Time Span Available: 1964 to date.

Frequency of Tape Issue: Weekly.

Average Number of Source Items Cited per Tape: 6500

Subscription Cost or Leasing Details:
Subscription is $20,000/year; the use is restricted under subscription contract.

Software Availability: Software packages are available from the source of the tapes and details will be supplied upon inquiry.

Additional Information: Institute for Scientific Information, 325 Chestnut Street, Philadelphia, Pennsylvania 19106, Attn: Melvin Weinstock. Telephone (215) 923-3300.

MARC (MACHINE-READABLE CATALOGING) DISTRIBUTION SERVICE

The Library of Congress, Washington, D.C.

Subject Matter: Current English language monographic cataloging data.

Type of Source Items Input: The entire input is taken from monographs which also includes government reports/documents and conference proceedings.

Subject Analysis/Indexing: An average of 1.5 subject headings are assigned each item. The items are classified using The Library of Congress Classification Schedules.

Searchable Data Elements: Each data element is searchable. Fixed and variable fields are included in the record.

Time Span Available: The past 2 years. 1969-1970.

Frequency of Tape Issue: Weekly

Average Number of Source Items Cited Per Tape: 1,200.

Subscription Cost or Leasing Details: Tapes are available on subscription for $800 per year.

Software Availability: None.

Type of In-House Service Offered: MARC tape searches are performed in-house on a time available and cost recovery basis.

Additional Information: The Library of Congress, Washington,D.C. 20540, Attn: Henriette D. Avram. Telephone (202) 426-6068.

INFORMATION ON MARC MAY BE FOUND IN THE FOLLOWING PUBLICATIONS:

The MARC Pilot Project: Final Report on a Project Sponsored by the Council on Library Resources, Inc. 1968. (183p.) For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20420, at $3.50 a copy. SC1.2:M18/2.

A detailed description of the format, character sets, bibliographic codes, input procedures, and cost of production of MARC records during the experimental MARC Pilot Project. Report includes summary descriptions of computer programs used and, in an appendix, reports written by each of the participating libraries that received and used the weekly magnetic tapes.

MARC Manuals Used by the Library of Congress; prepared by Information Systems Office, Library of Congress. 2d ed. Chicago ALA, ISAD, 1970. For sale by the American Library Association, 50 East Huron Street, Chicago, Illinois 60611, at $12.50 a copy.

This 4-part volume of more than 300 pages contains three handbooks and one special study. The first handbook "Books: A MARC Format," provides specifications for magnetic tapes in the MARC II format. It is designed for programmers who will be implementing systems using MARC records. The "Data Preparation Manual: MARC Editors." is a detailed guide to procedures followed by the MARC editors at The Library of Congress in preparing bibliographic records for conversion to machine-readable form. The "Transcription Manual" provides similar formation for the operators of the magnetic tape typewriters used in the MARC system. The special study, "Computer Magnetic Tape Usability Study," provides a list of data processing equipment which can handle MARC tapes.

BOOKS

A MARC Format: Specifications for Magnetic Tapes Containing Monographic Catalog Records in the MARC II Format. 4th ed. April 1970. (70 p.) Distribution free upon request to LC Card Division subscribers.

This volume is a revision of the document formerly published under the title: Subscriber's Guide to the MARC Distribution Service. It describes the tape and record formats, the character set, and the data fields in the MARC records.

Serials: A MARC Format. Washington, Library of Congress, 1970. (72p.). For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402, at $.70 a copy. LC1.2:M18/7.

Maps: A MARC Format. Washington, Library of Congress, 1970 (45p.) For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402, at $.05 a copy. LC1.2:M18/6.

Motion Pictures. Filmstrips, and Projected Meida Intended for Projection: A MARC Format, Washington, Library of Congress, (in preparation).

CONCLUSION

It is interesting to note that many of the information collections are available in both microform and magnetic tape form. Which is selected depends on the end use, as well as on the availability of computers, etc.

Most users of the listed services and other similar services agree that internal operations are enhanced considerably from viewpoints of time, money, manpower and convenience by using these "packages." However, a caution must be noted, that in the construction of these packages a generalization has been built in to make these acceptable to a relatively wide audience of subscribers. This generalization can prove to be disconcerting for people that have always done things in a precise fashion uniquely for their own organization needs.

In brief, the "package" concept of information services can be of great assistance to getting started rapidly and/or expanding existing activities with wide yet directed coverage.

REFERENCES

Carroll, K.D., Survey of Scientific-Technical Tape Services. New York, New York, American Institute of Physics. September 1970.

Dataflow Systems Incorporated, Master Index to Engineering Data Files in Sixteen Millimeter Microfilm Formats. 1st ed. Dover, New Jersey, Picatinny Arsenal. Sept.1970

Session Eight - Discussions

COMMERCIALLY AVAILABLE SERVICES

Chairman: Mr. D.G. Kingwill (South Africa)

DR. E. GARFIELD (US): The fact that beginning with January 1972 there will be a weekly subject index to Current Contents will have a very significant impact on computer services. One of the critical problems that computer oriented services have is timing. Computer services have become popular because in their own rather costly way, they have overcome time barriers which stood in the way of printed services. Whenever traditional services become competitive in time and cost, and certainly they are competitive costwise, the computer service will either have to improve in terms of timing or in terms of cost.

We are the largest producer of a magnetic tape base in the world. We cover approximately 400,000 articles per year. Many centers throughout the world, however, do not use our services. The main deterrent to greater use of our data base, which is international, multi-disciplinary, and complete in the coverage of 25,000 journals, is, among other things, that additional journal coverage is desirable.

The data base that we have constructed is based upon an analysis of citations which shows those journals which are most frequently cited in the scientific literature. I will soon be publishing a list of 500 most important journals of science. I am leaving with Mr. Keren in Israel a copy of a list of journals which I presented a few weeks ago at the Cranfield Conference in England. This list is important to show what the usual patterns of literature are. But although these 500 journals account for about 75% of all relevant literature, I do not underestimate the importance of additional hundreds and thousands of journals, that people want covered in specific situations.

Dr. Brown of the National Research Council of Canada gets from us a data base covering 25,000 journals and I am sure he can justify the need to cover 500 or 1000 more in various areas that are of interest to his users, and similarly Mr. Skov and others from Scandinavia. We know that the group at the Royal Institute of Technology in Sweden are processing journals on their own, to satisfy certain special needs.

If you can't justify the cost of importing journals on your own, then it is purely a matter of economics to come to ISI or any other data base supplier and say: here is a list of journals that we want you to add and here is the money to pay for it.

We do what is economically feasible but we cannot perform miracles. If we had large Government subsidies we could do more. I might point out that almost all the data bases that we are talking about today could not be offered at the prices at which they are offered were it not for the fact that they are by-products of other services we are producing like Science Citation Index, etc. After all, it is inconceivable that a data base which costs millions of dollars per year to produce, as ours does, could be sold for $8,000 to a dozen or more users.

As producers, we are ready and able at any time, whether through UNISIST or just informal discussions with our present or potential users, to add to this file any number of journals they want.

VIII Commercially Available Services

DR. J. BROWN (Canada): My paper was written several months ago. The service in Canada is developing rather rapidly, so that techniques have since changed somewhat and the figures stated in the paper have also changed.

The National Science Library operates as a division of the National Research Council of Canada. The name "Library" is misleading in this case, particularly if you interpret the word library in a traditional manner.

The National Science Library is really an information transferral agency whose job is to develop techniques to provide scientists, engineers and research workers with information on publications which they need in their day to day work.

The National Science Library originated as a National Research Council Library, but since its inception, it has been operated as a National Science Library. Its resources have been developed in very close collaboration with other major libraries in the country, so that the resources and the services provided by the National Science Library complement and supplement resources and services of other libraries.

The SDI service is regarded as an extension of the information services provided by the National Science Library. It is a very small part of it.

The SDI service, as it is now, operated on an experimental basis for about three years. Before being offered as a National Service we did a market analysis and attempted to find out of what use to the population it might be, what the interests of the users are, and what price they would be willing to pay for such a service.

We are now using 8 tape services: Chemical Titles; CA Condensates; ISI Source and Citation Data Tape; INSPEC; Biological Abstracts Previews; MEDLARS; GEO. REF; and MARC Distribution Service. Recently we have added the ERICTAPES, the COMPENDEX, and Metals Abstracts Index tapes. That makes 11 tapes.

We discovered very early that these tapes are incompatible as to software and hardware, so that one of our big problems was to develop a data base which would eliminate the necessity to rewrite programmes every time we added a new data base.

A great deal of time, effort and money was spent in developing what we call a standard file format, which enables us to take any data base and reprogram it to this standard.

The other problem we were faced with was the development of interest profiles. Here again we discovered early in the game that it is very difficult to develop interest profiles by telephone, by correspondence and so on. What is really required is somebody who is familiar with the techniques of SDI, who has a subject background and who can sit down with the user and develop his interest profile.

So we started a training program for what we call search editors. We offered a two day seminar in Ottawa to teach search editors how to use the SDI program. There was a great interest in this and people were willing to come to Ottawa at their own expense.

To date we have trained about 250 people located in university libraries, company libraries, industrial libraries, Industrial Research Council libraries, Government libraries, etc. Training of search editors is a crucial matter in this program and we are inclined to extend this further, and to train them on a more formal basis. They will be trained and paid by the National Science Library, and then located in specific regions across Canada, where their job will be to conduct local seminars, assist in the development of interest profiles, and travel around the country and sell the service.

We have found that some of the best search editors are good reference librarians. The techniques in constructing interest profiles, whether for mechanized or conventional systems, are the same.

When the service was first offered on a national scale, over two years ago, we did not have good cost figures. The price we charged initially was $100 per year, plus an additional $15 if the person wanted to use more than one data base. We recently have changed these prices because we now have a very good picture of what it is costing us. We find that the cost at the present time is $2 to $2.50 per profile, per search. The new price schedule represents our total cost; that is, computer time, personnel, paper, mailing charges and so on. It does not include the costs for leasing the tape services we are using.

We reasoned that as when compiling a bibliography, we do not charge the user for the cost of the literature used, we should not charge the user for the cost of acquiring tape services.

After the service had been in operation for a little over a year, we attempted to evaluate it. At present we have about 800 subscribers serving about 2000 people and we circulated a questionnaire asking which data bases they would like to have added. We also attempted to find out whether this service was saving the user any time.

We came up with rather astounding figures. People who answered the question were saving about 39,000 manhours per year. Converted into dollars, this is quite a substantial sum for relatively few people.

Some of the users even indicated that they no longer used conventional search techniques which we were not too happy about. People should continue to use normal procedures for literature searches and our service as a supplement to catch material that might have been missed.

One of the repercussions of our service is that the demand for literature loans has skyrocketed. When we initiated the service we promised subscribers that the National Science Library would hold all the journals which are cited. There is nothing more frustrating for a research worker receiving a list of citations than to find that he cannot get the original papers.

The figures I have given here represent the use of a service which has not really been publicized. We were operating under austerity conditions so we had to be rather cautious, but we are now taking steps to expand. In the survey of user reactions we found that there was a great demand for retrospective searches.

The only data bases which are available today to carry out retrospective searches are MEDLARS and the ISI Citation Tapes. A current awareness search is quite a different matter from a retrospective one. You cannot use the same program. You can run the tapes sequentially, but when you have a tape coming out monthly and you have five years accumulation, searching sixty tapes is a very costly business as far as computer time is concerned.

So we are struggling to develop a technique whereby we can do retrospective searches at a reasonable cost.

MR. H. SKOV (Denmark): I am going to represent the users point of view and report on the work done in Scandinavia regarding the real cost of documentation services based on magnetic tapes.

We have about 5 years experience in running services of that kind and we found it advisable to make a common pricing policy. But to do this we had to know how much it actually costs to run a tape service.

VIII Commercially Available Services

CA Condensates were chosen for analysis with fifty or alternately three hundred annual search profiles. Based on practical experience we assumed that the cost of a service to industry and research can be divided into four components: (i) disbursement to the supplier of the tape system; (ii) working expenses of the documentation center; (iii) cost of electronic data processing; (iv) expenses connected with the procurement of original material to the subscribers.

We organized four working groups. The first had to find out what would be the cost of a system operating for either fifty or three hundred profiles. The second had to find out the cost of all the practical work of the documentation center; the organizing of the system, selling the profiles, instructing people, and so on. The third dealt with electronic processing, which can be done by having your own computer or by buying time. The fourth investigated the cost of material per profile. The findings are stated in my paper.

The cost of one profile is about $800 where 50 profiles are served, and about $700 for 300 profiles. Costs will go up. Costs of documentation work depend on the efficiency of the service. To serve users well you must expect rising costs. Better programs, faster computers might make it easier in the future, but since the data base constantly increases, more machine time will be required.

The procurement of literature is indeed an increasingly expensive business. Supplying the customers with a copy of the original article requires time; we often have to write to remote cities, we have mailing expenses, maintenance of mailing lists, etc.

Our figures are a valuable aid to policy makers who decide whether or not to pay for a service. In Copenhagen we buy electronic processing time from a big computer center. We intend to cooperate, as far as possible, internationally, with other centers in Europe. I expect in the future that we shall exchange or buy profiles.

There are now about forty to fifty tape systems in the world. From some of the systems we could serve only one or two profiles in Denmark. That would not pay, so we buy them. If our figures are correct, a profile in chemicals costs about $800. Three of these profiles cost about the same as a subscription to the printed periodical. Does it pay for a firm to pay for a subscription or to use a few profiles? We tried to see it from the users point of view and we have made this survey to influence to some extent the data base manufacturers and also to give some figures to our financial authorities. Every country will have its own policy regarding how much shall be subsidized by the State and how much shall be paid by customers. The figures presented here are based on Scandinavian prices, but they have some use for other countries as well in their attempts to come to a sensible pricing policy.

MR. KINGWILL: The question is how many subscribers should a national agency have to make it worthwhile to subscribe to one of the commercial services and provide a service on a national basis as is done in several countries now.

The other aspect which concerns smaller countries particularly is the question of making the conventional sources available. When you contemplate subscribing to these tape services and providing the SDI service based on them, you are faced with the problem of providing the hard copy as well.

In South Africa this relates to the problem of building up scientific and technical information sources. The commercial services aim, as Dr. Garfield has pointed out, at providing information from the most important journals to the greatest number of users.

In his paper, Mr. Doudnikoff says that the medium versus the discipline was considered. The medium won out. Therefore, if discipline is of key interest, selections must be made from each medium. Microform, magnetic tapes, and hard copy publications are the media. For the purpose of his description the hard copy services are being de-emphasized. However, I want to bring to your attention the results of a survey made by one of my staff concerning the problem of the hard copy.

She found that the tape of the American Geological Institute (GEO. REF) enters items from about 1000 journals. Of these, 300 have all articles entered. SPIN tapes of the AIP dealing with all areas of physics and astronomy enters articles from 65 journals. Mathematical Offprint Service enters all articles from about 170 journals. Mathematical Reviews reviews 1,100 journals of which 180 are entered cover to cover. Basic Journal Abstracts of Chemical Abstracts enters articles from 35 leading journals in chemistry and chemical engineering. Chemical Titles processes 700 and the CITE tape of Engineering Index covers 300 "core" technological journals in plastics and electronics engineering.

This gives you some idea of the problem we are faced with.

MR. F. LIEBESNY (UK): To Dr. Garfield. I was rather distressed by your forecast that the costs are going up. I should have thought that in an economic organization, an efficient organization, the larger the turnover, the lower the cost. Why not?

DR. GARFIELD: To Mr. Liebesny. I do not know where you got the impression that I said the prices are going up. On the contrary, what I said was that if the size of the data base is to be increased beyond what it already is, we will be glad to enter into in any kind of cooperative arrangements with the tape suppliers whereby we can add to the already existing coverage of 2500 journals, if they want to help support the additional cost of input. I did not say that ISI had stopped adding journals.

We operate the ASCA system. Anyone who wants to, can subscribe tomorrow for about $135 a year in Israel, $115 in the United States. Incidentally, Mr. Skov might compare that to his cost of $800. I do not know the details of what he is searching for $800, but it seems to me that since we have a range of profile users beginning at $100 and going up to $2000 per year, that our average customer spends about $150 per year per profile. I am confident that we can provide weekly turnouts where the customer can sit at his desk and throw his own computer out.

In the first year of operation, we covered 600 journals. Without increasing the price during six years, we have increased the coverage to 2500 journals. That, in spite of inflation, is not an increase, but a very significant decrease in price.

There is no price increase contemplated for ASCA in the coming year. On the contrary, ISI tapes will certainly not go up, and may even go down. So the number of users of our tapes definitely affects the future and present pricing. Every time we get a new user, whether it is at $150 a year or $2000 a year, there is an allocation made for new journals. Whenever we get a substantial new user, let us say a subscription for tapes from somebody interested in petroleum, we would stipulate the coverage of X number of petroleum journals for that particular user.

MR. H. SCHUR (UK): To Dr. Brown and Mr. Skov. Could you give us some details of who your users are? Where do they work? Are they at universities, or in industry, and if in industry, are they in the large firms or in the small ones? I am particularly interested because in Denmark and in Canada the technical information service is something on the lines of the industrial liaison officers in the UK. I would, therefore, like to know whether any computerized services are also used by the TIS services and not just by the universities.

VIII Commercially Available Services

DR. BROWN: To Mr. Schur. The SDI service operated by the National Science Library is a very personalized service. At present, users from industry comprise about 20% of the total user population. This figure is going up. The data bases that were available at the time we started our service were heavily oriented towards the basic sciences. As we are now gradually adding other data bases, such as COMPENDEX, INSPEC, and so on, we are acquiring more and more users from industry.

TIS - Technical Information Service - is operated by the National Research Council and works in close cooperation with the National Science Library. TIS has also developed an SDI service based on the programs developed by the National Science Library. It is a less sophisticated system because it is directed towards an individual industry, a textile or furniture manufacturing industry, etc.

I suspect that many of the individual users of the NSL SDI service are also profiting by the SDI services offered by TIS.

An amusing thing came up in connection with cost. We have a very sophisticated system which searches the contents of about 9000 journals at a fantastic speed, over weekends, and the print-outs are sent out every Monday morning. We ran into a bottleneck because we had great difficulty in speeding up the mailing of these print-outs. We had to get a rather large clerical staff to tear the sheets apart and stuff them in envelopes. However, the problem is not yet solved.

MR. SKOV: To Mr. Schur. There is an approximate 50-50 distribution between research people and people from industry. Our system in Denmark hinders researchers from using more profiles because they have very small local budgets for purchasing information. In England students in their last year can get profiles without fee to get accustomed to this new system. Unfortunately we in Denmark cannot afford that for the time being.

MR. D.H. BARLOW (UK): To Mr. Skov. I think one of the main points that came out in your remarks, was the ratio between printed publications and tape services. Tape services are generally a spin off. As the number of centers using tapes increase, so will the provision of information to the various users, who are traditionally users of the printed publications. I think there is a very real problem that has to be faced by the users of information, that the producers have to maintain the balance between information produced through retail outlets and through wholesale. This has not been taken into account in previous studies.

MR. SKOV: To Mr. Barlow. As to the modern systems as against traditional ones, we have had some customers who have cancelled their subscriptions for quite a lot of periodicals. Maybe we killed the basis we work on. I do not know what will happen. The new system has great appeal to some people and many who never did very much conventional reading, do read now.

DR. GARFIELD: Mr. Kingwill mentioned the difficulty in deciding when a center is ready for tapes. How many profiles does this require? There are many centers which can justify tapes for a very small number of profiles for reasons of confidentiality. We have centers using our tapes who may have as few as 30 or 40 users. They do not want to buy a commercial service because they would have to submit their profiles which they prefer not to do. It is true that by examining profiles you can find out quite a bit about what the firm is doing. Some military organization might not particularly relish the idea. Also, the size of the individual profiles is very significant.

The final point regards the implementation of tape services. Every installation that we make, unless the user is sophisticated enough, or prefers not to have it, is supported by a training program. We do not sell any service that is not backed up by the proper installation and continuing training. Even the customer of a $150 profile gets help in the construction of his profile.

I might add that I spend more than fifty percent of my time on the problem of the so-called less developed countries. I do not consider Israel one of these countries, but in some South American or African countries, I would never recommend the installation of a tape system. They can, through our own commercially available services, buy, for $150 per profile, something which will be much easier for them to use and thus not be confronted with all the problems of programing, training computer people, training information scientists, etc. On the other hand, if there are political or other reasons why they must have a computerized service, we will send key people to those places to help them install those systems.

SESSION NINE

processing for automation-1

Authors of Papers

EVENT INTERACTION ANALYSIS OF ISRAEL AND OTHER MIDDLE EAST COUNTRIES USING THE FACILITIES AND DATA OF THE WORLD EVENT/INTERACTION SURVEY (WEIS)

Emil Schafer
University Computing Center
University of Southern California

SYNOPSIS

The use of four generalized programming languages to collect, structure, and analyze data related to political and other events or interactions among countries of the world is described. The data management language GIS is used to construct and maintain hierarchic files from which data is conditionally selected for statistical analysis using BI-MED and SPSS. Event data for Israel, UAR, Jordan, Syria, Lebonan, and Iraq are analyzed for the 4-year period 1966-69; chronology of events for June 1967 is listed, and frequency distributions of acts by country are plotted for the same month.

Introduction

The study of international political systems has become a major area of research in the field of international politics. This paper is a report of a project to incorporate such data into files capable of being conditionally queried for quantitative data analysis and qualitative review of event interactions. The relative ease and power of high-level data management languages together with packaged statistical programs are illustrated. These methods and techniques have been taught to classes of graduate students who have been able to use them to describe and construct data bases for individual and group research projects. In this way, students and other researchers with no prior exposure to data processing have been able to do fairly sophisticated data manipulation and to effectively increase their time available for interpretive study.

The World Event/Interaction Survey (WEIS) is a program of education and research in the Department of International Relations at the University of Southern California whose purpose is to develop the ability to describe, trace, compare, explain, and hopefully, to predict, under theoretical discipline, a particular range of international behavior by means of indicator data. The research is expected to contribute findings that might shed light on external performance traits and tendencies of countries, and also on configurations of actions and responses in the political exchanges among countries. The central topics are: how are problems coped with, and how are conflicts generated and resolved.

WEIS is also a data gathering and data analysis program to analyze structures and processes of actions and responses at what is defined as the "national control level" of international politics. Each data item is defined as a report of an international event, e.g.,

ITALY EXTENDED A TWO MILLION DOLLAR LOAN TO MOROCCO.

SYRIA REPORTED THAT SYRIAN PLANES SHOT DOWN 5 ISRAELI JETS.

There are almost 35,000 such data items presently in the WEIS files. These have been extracted from the descriptive and explanatory materials of news stories contained in daily newspapers. The problems, issues, contests, conflicts, and adjustments which arise in international political relations are the subject of WEIS data gathering and analysis. Each such data item is defined as an event/interaction. These are included as long as they relate to specific and discrete acts which have been committed. Excluded are editorial reports, speculations on future international affairs, and generalized reports on series of events. The WEIS is primarily interested in what nations do as they act and respond in international political situations.

Source Data Descriptions

Two source tape files were received for this demonstration project: the WEIS Analytic file and the WEIS Descriptive file. The Analytic file is a file whose fields are compatible with access by several statistical programs; the Descriptive file contains short abstracts or descriptions of the event interactions. For each record in the Analytic file, there should be a corresponding record in the Descriptive file. These two source files were constructed of 80-column (or 80-byte) card images containing the fields shown in Table 1.

ANALYTIC FILE FIELDS	DESCRIPTIVE FILE FIELDS
YEAR	MONTH
MONTH	DAY
DAY	YEAR
ACTOR	DESCRIPT
EVENT	COLLID
TARGET	SERIALNO
SOURCE	LINEID
CODER	DECKTYPE
ARENA	
COLLID	
SERIALNO	
LINEID	
DECKTYPE	

Table 1. Analytic and Descriptive Source File Fields.

The fields ACTOR and TARGET contain numerical codes which designate countries initiating and receiving some action; the type of action is coded in EVENT. The field ARENA localizes the area in which the event took place, while SOURCE codes the newspaper source of the data. The collection identification is encoded in COLLID, while the individual coding the raw data is designated in CODER. The field LINEID has meaning only for the Descriptive file where several source cards may constitute one logical record. DECKTYPE identifies the file: a "1" is the Analytic file; a "2" is the Descriptive file. The fields YEAR, MONTH, DAY, and SERIAL NO are self-evident. (Details of the coding used for both files are contained in Ref. 1). Figure 1, below, contains several examples of the types of coding used.

ACTOR or TARGET	EVENT/INTERACTIONS	SOURCE
002 USA	011 Surrender; yield to order; submit to arrest.	01 New York Times
070 Mexico	012 Yield position; retreat; evacuate.	05 Middle East Journal Chronologies
305 Austria	013 Admit wrongdoing; retract statement.	12 Times of India
651 UAR		17 Middle East Journal
666 Israel		

COLLID	ARENA
02 Berlin crisis	010 Arab-Israeli conflict (general)
17 Middle East	013 Events during 1967
12 WEIS 1968	020 Vietnam conflict
24 Sino-Soviet interactions	050 Sino-Soviet conflict
	160 Red Guard activities

Fig. 1. Representative code and decode values for Analytic and Descriptive fields.

An example showing the same record, as it might be coded for both the Analytic and Descriptive formats is shown in Fig. 2.

DESCRIPTIVE

```
10 17 68 VTN REJECTS USA DEMAND FOR RECIPROCITY FOR  12064212
10 17 68 BOMBINGS OF VTN.                            12064222
```

ANALYTIC

```
68 10 17 816 111 002 01 26 025                       12064201
```

Fig. 2. Source Analytic and Descriptive card images.

Creating the Analytic Data Set

To create the new Analytic data set, two file descriptions were written using the GIS data description language. The first DDT (or Data Description Table) describes the source file; the second, the new file to be created. The DDT for the source file is shown in Fig.3, while that for the new Analytic file is shown in Fig.4. The new Analytic file is described as a one-level GIS file, like the source file, except in two respects: (1) Each record is compressed in size (from 80 to 30 bytes) to remove intervening blanks, and (2) The file is defined to be a generation data group (GDG) which makes it possible to use GIS for easy file maintenance functions. It was also desired to provide automatic field decoding facilities for all the coded fields. This can be done by including decode tables in the DDT, (see Fig.4). For example, if the coded value "09" is presented for the field MONTH, the decode function will print out "SEPTEMBER". Again, if the field ACTOR has the coded value "130", the decode function will generate "ECU", the acronym for the country Ecudor. The keyword ENTRIES=2 in the DATM statement indicates that GIS will maintain two generations of the file. The SEGM or segment statement serves to collect all the field statements above it into a GIS segment. The decode functions may be easily controlled by the user, i.e., they may be turned off when not wanted.

```
DDT;
FILE:NAME=WEISATAP;
FLD: NAME=YEAR,      LENGTH=2;
FLD: NAME=BLANK1,    LENGTH=1;
FLD: NAME=MONTH,     LENGTH=2;
FLD: NAME=BLANK2,    LENGTH=1;
FLD: NAME=DAY,       LENGTH=2;
FLD: NAME=BLANK3,    LENGTH=1;
FLD: NAME=ACTOR,     LENGTH=3;
FLD: NAME=BLANK4,    LENGTH=1;
FLD: NAME=EVENT,     LENGTH=3;
FLD: NAME=BLANK5,    LENGTH=1;
FLD: NAME=TARGET,    LENGTH=3;
FLD: NAME=BLANK6,    LENGTH=5;
FLD: NAME=SOURCE,    LENGTH=2;
FLD: NAME=BLANK7,    LENGTH=1;
FLD: NAME=CODER,     LENGTH=2;
FLD: NAME=BLANK8,    LENGTH=38;
FLD: NAME=ARENA,     LENGTH=3;
FLD: NAME=BLANK9,    LENGTH=1;
FLD: NAME=COLLID,    LENGTH=2;
FLD: NAME=SERIALNO,  LENGTH=4;
FLD: NAME=LINEID,    LENGTH=1;
FLD: NAME=DECKTYPE,  LENGTH=1;
SEGM:NAME=ANALYTIC, LEVEL=00, ;
#TYPE=RECORD, UNIND=Y, ;
#SORT=YEAR,A,SERIALNO,A;
*;
DATM:DSORG=PS, CREATE=NO, CATLG=;
#YES, UNIT=2400, DCB=(DEN=3), ;
#LRECL=80, BLKSIZE=3200, RECFM=;
#FB, VOLUME=SER=001006, ;
#DSNAME=WEISDATA;
END;
```

Fig.3. DDT for the Source Analytic File.

By providing additional DDT'S for the same file (the same DSNAME), it is possible to readily describe the file differently. For example, if one would want to have the field names print out in words of another language, a simple addition to the field statement makes this feasible.

The final step in the creation of the new Analytic file was to write a small program in the GIS procedural language to define the mapping of fields from one file to another. This mapping is usually partitioned on a segment basis. The conversion program, as shown in Fig.5, reads source file records into a special GIS file (HOLD5) where they may be sorted. The word QUERY initiates all the processes of locating the data set in the system catalog, requesting tape mounts, opening files, etc. The word SORT defines and initiates the sorting actions. Finally, the word CREATE initiates

```
DDT ;
* THIS DDT DEFINES THE WEIS ANALYTIC FILE;
FILE:NAME=WEISAGIS;
FLD: NAME=YEAR, LENGTH=2, JUST=L;
DECD:TYPSPC=LKUP, LGTHA=2, LGTHF=4, EEDVAL=66,1966,67,1967,68,1968,;
#69,1969,70,1970,71,1971,72,1972;
FLD: NAME=MONTH, LENGTH=2, JUST=L;
DECD:TYPSPC=LKUP, LGTHA=2, LGTHF=9, EEDVAL=;
#01, JANUARY ,;
#02, FEBRUARY,;
#03,  MARCH  ,;
#04,  APRIL  ,;
#05,   MAY   ,;
#06,  JUNE   ,;
#07,  JULY   ,;
#08, AUGUST  ,;
#09,SEPTEMBER,;
#10, OCTOBER ,;
#11,NOVEMBER ,;
#12,DECEMBER ;
FLD: NAME=DAY, LENGTH=2, JUST=L;
FLD: NAME=ACTOR, LENGTH=3, JUST=L;
DECD:TYPSPC=LKUP, LGTHA=3, LGTHF=3, CONVA=S, CONVF=S, ;
#FEDVAL=002,USA,020,CAN,040,CUB,041,HAI,042,DOM,051,JAM,052,TRI,053,;
#BAR,070,MEX,090,GUA,091,HON,092,ELS,093,NIC,094,COS,095,PAN,100,COL,;
#101,VEN,110,GUY,130,ECU,135,PER,140,BRA,145,BOL,150,PAR,155,CHL,160,;
[illegible]C,165,URU,198,AFR,[illegible]99,OAS,200,UNK,205,IRE,210,NTH,211,BEL,212,LUX,;
FL[illegible]N,221,MOC[illegible]225,SWZ,230,S[illegible]32,AND,235,[illegible]GMW,265,;
FLD: NA[illegible]E[illegible] LENG[illegible],305,AU[illegible] H[illegible]315,C[illegible]VAT,;
FLD: NAME=[illegible]KTYPE, LENGTH=[illegible]=L, HEA[illegible]TYPE;
*;
SEGM:NAME=ANALYTIC, LEVEL=00, TYPE=RECORD, UNIND=Y, ;
#SORT=YEAR,A,SERIALNO,A;
*;
DATA:DSORG=PS, CREATE=YES, CATLG=YES, DSNAME=H183ARU.WEISANTR,
ENTRIES=2, DCB=(DEN=3), LRECL=30, BLKSIZE=3000, RECFM=FB;
END;
```

Fig.4. DDT for the new Analytic file incorporating decode tables for all fields.

the creation of the new file as defined by the mapping of the STRUCTURE statement. The verb INSERT implies the insertion of the entire segment on a non-duplicating basis. The last step is not a part of the file conversion; it merely lists out the first 25 records of the new file for visual inspection. Fig.6 shows a brief listing of these records in an automatic GIS format provided from the information stored in the DDT. Note that the decode table for the field ARENA includes a print value of FIELD EMPTY for an empty field; otherwise GIS will indicate this condition by a string of asterisks. If there is no decode function for a given argument in the decode table, GIS will indicate this with a string of pound signs (####). As part of the file creation process, GIS performs the cataloging of the new data set.

```
QUERY WEISAGIS
01 YEAR  MONTH          DAY  ACTOR                                                   TARGET
02 SOURCE                         EVENT
03 SERIAL NO.  LINE ID  DECK TYPE  CODER  ARENA                    COLLECTION ID

01 1967    JUNE          01   USA   PROMISE OWN POLICY SUPPORT                        UNK
02 NEW YORK TIMES                        25         FIELD EMPTY     WEIS 1967
03 2269                         1

01 1967    JUNE          01   UAR   GIVE WARNING                                      NSC
02 NEW YORK TIMES                        25         FIELD EMPTY     WEIS 1967
03 2270                         1

01 1967    JUNE          01   USA   COMMENT ON SITUATION - NEUTRAL                    NSC
02 NEW YORK TIMES                        25         FIELD EMPTY     WEIS 1967
03 2271                         1
```

Fig.6. Listing of several records from new Analytic file in default automatic format using information from DDT.

Creating the Descriptive Data Set

In a manner parallel to that for the Analytic file discussed above, two DDT's were written for the Descriptive files. The first was to describe the source file; the second, to describe a GIS hierarchic file. The objective here was to create a file of varying record length due to a variable number of lines of abstract or description. This was accomplished by placing the lines of descriptive text in subordinate segments; subordinate, that is, to a master or zero-level segment containing record identification and other information. A count field in this segment keeps count of the number of lines of description in each record. The DDT for the new hierarchic file (Fig.7), describes a variable length, blocked record, containing a variable number (up to nine) of fixed length segments called SUB, each of which contains one line of description, and one fixed length segment named MSTR.

```
QUERY WEISATAP
LOCATE RECORD
HOLD HOLD5 RECORD
EXHAUST RECORD

SORT HOLD5 ASC YEAR, SERIALNO

CREATE WEISAGIS FROM HOLD5
STRUCTURE ANALYTIC FROM ANALYTIC
INSERT ANALYTIC

QUERY WEISAGIS(+1)
LOCATE RECORD
LIST RECORD
EXHAUST RECORD OR 25
END PROCEDURE
```

Fig.5. Three-step GIS procedure to create new Analytic data set. Fourth step only reads 25 records of new file.

```
DDT ;

FILF:NAMF=WEISDGIS;
FLD: NAME=YEAR, LENGTH=2, JUST=L;
DECD:TYPSPC=LKUP, LGTHA=2, LGTHF=4, EEDVAL=66,1966,67,1967,68,1968,;
#69,1969,70,1970,71,1971,72,1972;
FLD: NAMF=SFRIALNO, LENGTH=4,   JUST=L, HEADFP=SERIAL NO.;
FLD: NAMF=MONTH, LENGTH=2, JUST=L;
DECD:TYPSPC=LKUP, LGTHA=2, LGTHF=9, EEDVAL=;
#01, JANUARY ,;
     FEBRUARY
#12,D
FLD: NAME=DAY, LENGTH=2, JUST=L;
FLD: NAME=COLLID,   LENGTH=2,   JUST=L, HEADER=COLLECTION ID;
DECD:TYPSPC=LKUP, LGTHA=2, LGTHF=40, CONVA=S, CONVF=S, EEDVAL=;
#01,SINO-INDIAN BORDER WAR (HOGGARD)         ,;
#02,BERLIN CRISIS (MCCLELLAND)               ,;
#03,TAIWAN CRISIS (C   LAKE STUDY)           ,;
    VIETNAM WAR (    (ME    T)
#21,1     INTE    AL CONFLIC     TZSIMMONS),;
#22,16 INTE   TIONAL CONFLICTS (FITZSIMMONS),;
#23,16 INTERNATIONAL CONFLICTS (FITZSIMMONS),;
#24,SINO-INDIAN INTERACTIONS (HOGGARD)       ,;
#27,TIMES OF INDIA SAMPLE                     ;
FLD: NAMF=DESCCNT,  LENGTH=1,   JUST=R, HEADER=NUMBER DESC LINES;
*
SEGM:NAME=MSTR, LEVEL=00, TYPE=RECORD, UNIND=Y, SORT=YEAR,A,SFRIALNO,A
*
DATM:DSORG=PS, CREATE=YES, CATLG=YES, ENTRIES=3, DSNAMF=H183ARU.INTREL
#BLKSIZE=6524, LRECL=652, RECFM=VB, UNIT=2400, DCB=(DEN=3);
*
FLD: NAME=DESCRIPT, LENGTH=62, JUST=L, HEADER=DESCRIPTION;
FLD: NAME=LINEID,   LENGTH=1,   JUST=L, HEADER=LINE ID;
*
SEGM:NAME=SUB, LEVEL=01, TYPE=TRAILR, OPTION=CNT, ;
#OPTFNM=DESCCNT, UNIND=Y, SORT=LINEID,A; | END;
```

Fig.7. DDT for new hierarchic GIS Descriptive file indicating a 3-generation data set.

Like the DDT for the Analytic file, the DDT for the Descriptive file contains many decode tables; note the decode table for COLLID. The file conversion procedure is very similar to that for the Analytic file with one important exception: since there is now a subordinate segment, the mapping statement is "up" from the lowest level segment (the SUB segment). There is only one structure partition, STRUCTURE SUB FROM MSTR, where SUB is the name of the segment in the file to be created, and MSTR is the name of the source file segment (the source file only has one segment, i.e., MSTR). The entire procedure to create this two-level hierarchic file is shown in Fig.8.

```
QUERY WEISDSOR;
READ RECORD;
HOLD HOLD20 RECORD;
EXHAUST RECORD;

SORT HOLD20 ASC YEAR, ;
#SERIALNO, LINEID;

CREATE WEISDGIS FROM HOLD20;
STRUCTURE SUB FROM MSTR;
INSERT SUB;

QUERY WEISDGIS(+1);
READ RECORD;
LIST RECORD;
EXHAUST RECORD OR 25;
END PROCEDURE;
```

Fig.8. Three-step GIS procedure to create new hierarchic Descriptive data set. (Fourth step only reads first 25 records).

Information Retrieval with GIS

Using the two data sets just created, one might be interested in several kinds of "information retrieval". First, are the extraction of subsets from the Analytic file in a format amenable to further processing by one or more statistical programs. Second, one might wish to be able to read a chronology composed of abstracts from a certain time sequence, e.g., all events during May 1968. Third, one might wish to query both files using the Analytic file as an index or parameter list to retrieve particular records from the Descriptive file. Information retrieval for any of these three purposes is relatively easy using the GIS data management language. In the examples which follow, we will demonstrate information retrieval for each of the three types. (In these examples, processing was always in the batch mode; GIS has teleprocessing facilities but they were not implemented at the time these programs were run.)

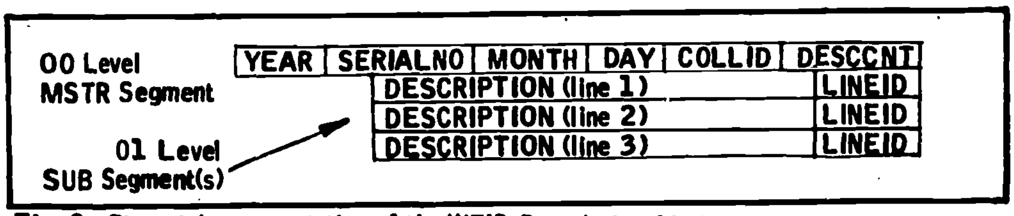


Fig.9. Pictorial representation of the WEIS Descriptive file hierarchic structure.

Action Within the Middle East Group, 1966-69

In this first example of information retrieval, we wish to construct a table showing frequency and percentage of event/interactions among the six countries ISR, UAR, SYR, JOR, LEB, and IRQ for the four year period 1966-69. That is, we wish to generate a subset from the WEIS Analytic file where the ACTOR and TARGET is any of these countries. The subset will then be processed by a statistical program to generate the required table. The first step is to write a DDT (a few minutes work), which describes the file (subset) to be created (see Fig.10). Note that this file includes only selected fields. Since the mapping defined by the STRUCTURE statement is by field name, within the named segment, only the named fields are mapped into the subset file. The second step was to write an SPSS (Statistical Package for the Social Science) program to construct the required table. (See Fig.11). The third step was the writing of a simple procedure to create the subset of the Analytic file.

```
DDT;
FILE:NAME=WEISDEM1;
FLD: NAME=YEAR, LENGTH=2;
FLD: NAME=MONTH, LENGTH=2;
FLD: NAME=DAY, LENGTH=2;
FLD: NAME=ACTOR, LENGTH=3;
FLD: NAME=EVENT, LENGTH=3;
FLD: NAME=TARGET, LENGTH=3;
SEGM:NAME=DEMO1, LEVEL=00, ;
#TYPE=RECORD, UNIND=N, SORT=;
#YEAR,A;
DATM:DSORG=PS, CREATE=YES, ;
#CATLG=YES, UNIT=2400, ;
#RECFM=FB, LRECL=15, BLKSIZE=;
#3000, DCB=(DEN=3), ;
#DSNAME=H183ARU.SUBSET1;
END;
```

```
QUERY WEISAGIS;
LOCATE RECORD;
IF ACTOR EQ '645','651','652',;
          #'660','663','666';
AND TARGET EQ '645','651','652',;
          #'660','663','666';
   HOLD HOLD1 RECORD;
EXHAUST RECORD;

SORT HOLD1 ASC YEAR, SERIALNO;

CREATE WEISDEM1 FROM HOLD1;
STRUCTURE DEMO FROM ANALYTIC;
APPEND DEMO;
END PROCEDURE;
```

Fig.10. DDT defining file for use by an SPSS program. The file creation procedure (on the right), selects records conditionally from the Analytic file.

```
// EXEC  SPSS360,TIME=3,REGION=150K,PARM='10000'
//GO.FT02F001  DD  DUMMY
//GO.FT08F001  DD  DSN=H183ARU.SUBSET1,DISP=OLD
//SYSIN  DD  *
  RUN NAME        WEISDEMO
  FILE NAME       WEISDATA
  VARIABLE LIST   YEAR,MONTH,DAY,ACTOR,EVENT,TARGET
  INPUT MEDIUM    TAPE
  # OF CASES      ESTIMATED 2000
  INPUT FORMAT    FIXED (3F2.0,F3.0,F2.0,1X,F3.0)

RECODE            ACTOR,TARGET (1 THRU 644,646 THRU 650,653 THRU 659,
                  661,662,664,665=0) (645=6)(651=2)(652=3)(660=5)
                  (663=4)(666=1) (667 THRU HIGHEST=0)
RECODE            YEAR(66=1)(67=2)(68=3)(69=4)
VALUE LABELS      ACTOR,TARGET (0) OTHER (1) ISR (2) UAR (3) SYR (4)
                  JOR (5) LEB (6) IRQ /YEAR (1) 1966 (2) 1967
                  (3) 1968 (4) 1969
PRINT FORMATS     YEAR TO TARGET (0)
FASTABS           VARIABLES=ACTOR(0,6)/YEAR(1,4)
                  TABLES=ACTOR BY YEAR
OPTIONS           3,5
READ INPUT DATA
```

Fig.11. SPSS program to process data in SUBSET1 file. The first RECODE statement assigns values to the ordinates, the second to the abcissa. VALUE LABELS assigns row and column labels; FASTABS dimensions the table array; title is assigned by TABLES= statement.

COUNT / COL PCT	YEAR 1966 (1)	1967 (2)	1968 (3)	1969 (4)	ROW TOTAL
ACTOR 1 ISR	54 35.1	125 39.2	158 43.6	299 47.1	636 43.3
2 UAR	26 16.9	69 21.6	68 18.8	186 29.3	349 23.7
3 SYR	40 26.0	54 16.9	16 4.4	35 5.5	145 9.9
4 JOR	23 14.9	55 17.2	103 28.5	72 11.3	253 17.2
5 LEB	1 0.6	3 0.9	14 3.9	26 4.1	44 3.0
6 IRQ	10 6.5	13 4.1	3 0.8	17 2.7	43 2.9
COLUMN TOTAL	154 10.5	319 21.7	362 24.6	635 43.2	1470 100.0

Fig.12. Table constructed by SPSS program of Fig.11 showing frequencies and percentages of event/interactions among the six countries ISR, UAR, SYR, JOR, LEB, and IRQ for the period 1966-69. Each element of the array contains absolute counts and column percentages. Total number of event/interactions is 1470.

This table illustrates information retrieval and organization using GIS and SPSS. Data in this table may be interpreted as follows: During 1966 ISR was the ACTOR country among the group 35.1% of the time with 54 out of 154 yearly events. During that year SYR was in second place with 40 events and 26.0% of the action. However, during 1967, second place went to UAR with 21.6% of the event action. In 1968, JOR overtook UAR for second place with 28.5% of the event activity.

The table in Fig.12 shows event/interactions on an annual basis. In order to prepare for some statistical programs, another table was prepared, like that in Fig.12, but indicating event/interactions on a monthly basis. A typical row in the table of crosstabulation totals by country would be:

	ISR	UAR	SYR	JOR	LEB	IRQ	ROW TOTAL
BY ACTOR	2	1	6	1	1	1	12

Using the data in the "Monthly by Actor" table, punched cards equivalent to each row of the table were prepared for each month in which the row total was not zero. This set of data was input to the BI-MED program BMD03D (Correlation with deletion) which computes a simple correlation matrix, omitting values of variables which the user may specify. The output from this program includes (1) Correlation matrix printed, and (2) Number of pairs of observations used in computing each correlation coefficient. Another similar program, BMD02D (Correlation with transgeneration) computes correlation coeficients, averages, and measures of dispersion on entering variables and/or transgenerated variables from selected cases whose values for specified variables have a precise logical relationship in agreement with a specified Boolean expression. Output from this program includes (1) Sums, (2) Means, (3) Cross-product deviations, (4) Standard deviations, (5) Variance-covariance matrix, and (6) Correlation matrix. Optional output includes a one-page cross-tabulation plot of any two variables.

Regression Analysis of GIS-Generated Data Set

This second example of information retrieval, data organization, and data analysis using GIS-created files and the SPSS statistical package, uses the same data set created for the previous example (See Fig.10). What is desired now, however, is to compare ISR as ACTOR to ISR as TARGET (ISR output to ISR intake) from three countries UAR, JOR, and SYR, for month-by-month distributions. For this data, multiple and stepwise regression regressions were

```
//  EXEC  SPSS360,TIME=4,REGION=200K,PARM='40000'
RUN NAME          WEISDEMO--REGRESSION ANALYSES
VARIABLE LIST     ISROUT FROMUAR FROMJOR FROMSYR
SUBFILE LIST      PROB4B,PROB5B,PROB6B
# OF CASES        48  48  48
PROCESS SBFILES   EACH
INPUT FORMAT      FIXED (4F3.0)
REGRESSION        VARIABLES=ISROUT TO FROMSYR
                  REGRESSION=ISROUT WITH FROMUAR TO FROMSYR (2)/
                  REGRESSION=ISROUT WITH FROMUAR TO FROMSYR (1)
READ INPUT DATA
```

Fig.13. Program deck for SPSS regression analysis. PROCESS SBFILES EACH indicates that each subfile is to be considered separately and in order. Each subfile consists of 48 cases (months); subfile 4B compares all events; subfile 5B calculates for all "cooperative" events; while subfile 6B calculates for "conflict" events. The first line of the REGRESSION statement requests multiple regression; the second, stepwise regression.

done with intakes from ISR from each of the three countries as independent variables, and ISR output as dependent variable. This analysis was performed for three types of event/interaction: all events; "cooperative" events; and "conflict" events. For illustrative purposes, only the analysis for cooperative events will be shown here. The SPSS program for these regression analyses is given in Fig.13. The output from this program, multiple stepwise regression calculations for cooperative events in shown in Fig.14. In addition to these SPSS calculations,

```
* * * * * * * * * * * * * * * * * * * * * * * * * M U L T I P L E   R E G R E S S I O N * * * * * * * * * * * * * * * * * * * * * * * * *

DEPENDENT VARIABLE..    ISROUT

VARIABLE(S) ENTERED ON STEP NUMBER  1..    FROMUAR

MULTIPLE R          0.68531                 ANALYSIS OF VARIANCE   DF      SUM OF SQUARES      MEAN SQUARE        F
R SQUARE            0.46965                 REGRESSION             1.          43.79509          43.79509     40.73591
STANDARD ERROR      1.03687                 RESIDUAL              46.          48.45495           1.07511

---------------- VARIABLES IN THE EQUATION -----------------      ------------- VARIABLES NOT IN THE EQUATION -------------

VARIABLE              B          BETA     STD ERROR B      F        VARIABLE      BETA IN     PARTIAL   TOLERANCE        F

FROMUAR           1.07473     0.68531      0.16839     40.736       FROMJOR       0.07249     0.00951    0.99911     0.491
(CONSTANT)        0.30242                                           FROMSYR       0.29280     0.38674    0.99190     7.914
VARIABLE(S) ENTERED ON STEP NUMBER  2..    FROMSYR

MULTIPLE R          0.74093                 ANALYSIS OF VARIANCE   DF      SUM OF SQUARES      MEAN SQUARE        F
R SQUARE            0.54897                 REGRESSION             2.          51.19182          25.59591     27.38625
STANDARD ERROR      0.96678                 RESIDUAL              45.          42.05818           0.93463

---------------- VARIABLES IN THE EQUATION -----------------      ------------- VARIABLES NOT IN THE EQUATION -------------

VARIABLE              B          BETA     STD ERROR B      F        VARIABLE      BETA IN     PARTIAL   TOLERANCE        F

FROMUAR           1.03457     0.65971      0.15765     43.087       FROMJOR       0.07892     0.11745    0.99310     0.615
FROMSYR           0.91195     0.29280      0.28862      7.914
(CONSTANT)        0.24058
VARIABLE(S) ENTERED ON STEP NUMBER  3..    FROMJOR

MULTIPLE R          0.74511                 ANALYSIS OF VARIANCE   DF      SUM OF SQUARES      MEAN SQUARE        F
R SQUARE            0.55520                 REGRESSION             3.          51.77197          17.25732     18.37541
STANDARD ERROR      0.97092                 RESIDUAL              44.          41.47803           0.94268

---------------- VARIABLES IN THE EQUATION -----------------      ------------- VARIABLES NOT IN THE EQUATION -------------

VARIABLE              B          BETA     STD ERROR B      F        VARIABLE      BETA IN     PARTIAL   TOLERANCE        F

FROMUAR           1.03107     0.65747      0.15839     42.376
FROMSYR           0.91711     0.29468      0.28996      7.942
FROMJOR           0.11834     0.07892      0.15084      0.615
(CONSTANT)        0.21717
                       SUMMARY TABLE

VARIABLE      MULTIPLE R   R SQUARE   RSQ CHANGE    SIMPLE R              B               BETA

FROMUAR         0.68531     0.46965     0.46965     0.68531           1.03107          0.65747
FROMSYR         0.74093     0.54897     0.07932     0.36253           0.91711          0.29468
FROMJOR         0.74511     0.55520     0.00622     0.09044           0.11834          0.07892
(CONSTANT)                                                            0.21717
```

Fig.14. Stepwise multiple regression calculations by SPSS program comparing ISR output to ISR intake from UAR, JOR, and SYR for cooperative events.

similar calculations were carried out using the BMD03R program (Multiple regression with case combinations). Output from this program includes (1) Sums and sums of squares, (2) Regression coefficients with standard errors and t-values, and deviation about regression with degrees of freedom and F-values, (3) Partial correlation coefficients, and (4) Intercept value. Also, for each selection and subproblem, a table of residuals can be obtained. Illustrated in Fig.15 is the output from the BI-MED multiple regression calculations for cooperative event/interactions data.

```
SELECTION NO.    2- 2

SAMPLE SIZE       48
NO. OF VARIABLES  4      NO. OF VARIABLES DELETED  0
DEPENDENT VARIABLE IS NOW NO.  1

COEFFICIENT OF DETERMINATION  0.5552
MULTIPLE CORR. COEFFICIENT    0.7451

SUM OF SQUARES ATTRIBUTABLE TO REGRESSION        51.77179
SUM OF SQUARES OF DEVIATION FROM REGRESSION      41.47821

VARIANCE OF ESTIMATE        0.94269
STD. ERROR OF ESTIMATE      0.97092

INTERCEP. (A VALUE)     0.21717


               ANALYSIS OF VARIANCE FOR THE MULTIPLE
                        LINEAR  REGRESSION
      SOURCE OF VARIATION        D.F.      SUM OF        MEAN        F
                                           SQUARES       SQUARES     VALUE
DUE TO REGRESSION.............    3        51.77179      17.25726    18.3065
DEVIATION ABOUT REGRESSION....   44        41.47821       0.94269
                      TOTAL...   47        93.25000


VARIABLE     MEAN        STD.       REG.     STD.ERROR    COMPUTED    PARTIAL      SUM OF SQ.  PROP. VAR.
  NO.                  DEVIATION   COEFF.   OF REG.COE.   T VALUE    CORR. COE.     ADDED        CUM.
   2       0.45833     0.89819    1.03106    0.15839     6.50965     0.70043     43.79504     0.46965
   3       0.27083     0.93943    0.11834    0.15084     0.78448     0.11745      0.48973     0.00525
   4       0.18750     0.49060    0.81711    0.28994     2.81822     0.39104      7.48718     0.08029
   1       0.87500     1.40896

COMP. CHECK ON FINAL COEFF.        0.81711


VARIABLES DELETED...  0
```

Fig.15. Multiple regression calculations output by BMD03R program, comparing ISR output to ISR intakes from UAR, JOR, and SYR for cooperative event/interactions in the period 1966 - 1969.

Information Retrieval from both the Analytic and Descriptive Data Sets

For this example, we wish to select and print out abstracts from the Descriptive file relating to records in the Analytic file where the ACTOR was USA, USR, UAR, or ISR, and the time was June 1967. What has to be done in effect is to search the Analytic file until a record is found satisfying these conditions, then run through the Descriptive file until the corresponding record is found. Then data from fields in both files will be printed out.

The file search logic portion of the GIS procedure to select and print out the required abstracts is given in Fig.16. The QUERY statement indicates the two files to be searched. The Analytic file (WEISAGIS) is searched to locate a record meeting the required conditions. The numbers in the ACTOR EQ statement are the code numbers corresponding to the four countries: 002=USA, 365=USR, 651=UAR, and 666=ISR. The READ WEISDGIS statement indicates that the descriptive file is to be read until a record IDNUMBER in that file is found which corresponds to that previously located in the Analytic file (YEAR and SERIALNO of the Analytic file concatenated). When such a record is found, the subordinate segments containing the lines of abstract or description are to be read until the subordinate (SUB) segments are exhausted (EXHAUST SUB). Since only one such Descriptive record is to be read for each

```
QUERY WEISAGIS, WEISDGIS
LOCATE WEISAGIS:RECORD
  WHEN YEAR EQ '67'
  AND MONTH EQ '06'
  AND ACTOR EQ '002','365','651','666'
READ WEISDGIS:RECORD
  WHEN IDNUMBER EQ WEISAGIS:YEAR|WEISAGIS:SERIALNO
    READ SUB

   | Report Format Statements |

    EXHAUST SUB;
EXHAUST WEISDGIS:RECORD OR 1
EXHAUST WEISAGIS:RECORD
END PROCEDURE
```

Fig.16. Logic search portion of program to search both Analytic and Descriptive files for records of event/interactions during June 1967 where the ACTOR was USA ,USR ,UAR or ISR.

Analytic record, the program drops down to the next statement, EXHAUST WEISDGIS: RECORD OR 1. Control is then given to the next statement, which continues the search of the Analytic file until all qualifying records have been LOCATE'd. The statements omitted in the partial program shown in Fig.16 serve to control the format and location of data from the fields of both files, to control lines per page, and to qualify the repetition of a field value. For example, one would not wish the value of the field DAY to print out for every record; rather, it is desirable to print out the value only when the value changes. This is readily done by the qualifying statement: DAY ON DAY CH. This prints DAY only when the value of DAY CHanges. A sample page output from this search is shown in Fig.17.

```
                                CHRONOLOGY FOR JUNE 1967

DAY  SERIAL NO  ACTOR  TARGET                           DESCRIPTION


 09     2519     UAR    USR           NEW YORK TIMES                          FIELD EMPTY
                                        NON-MILITARY DEMONSTRATION, WALK-OUT ON
                                      UAR DEMONSTRATORS TRIED THREE TIMES TO STORM USR EMBASSY IN UA


        2522     UAR    ISR           NEW YORK TIMES                     ARAB-ISRAELI 1967 WAR
                                        SUSPEND NEGATIVE SANCTIONS, TRUCE
                                      UAR ANNOUNCES WAR WITH ISR ENDED


        2523     ISR    UAR           NEW YORK TIMES                     ARAB-ISRAELI 1967 WAR
                                        SUSPEND NEGATIVE SANCTIONS, TRUCE
                                      UAR ANNOUNCES WAR WITH ISR ENDED


        2525     ISR    USA           NEW YORK TIMES                     ARAB-ISRAELI 1967 WAR
                                        CHARGE, CRITICIZE, BLAME, DISAPPROVE
                                      ISR PRM SAID USA PRS HAD PROMISED GREAT THINGS BUT IN END ISR
                                      STOOD VIRTUALLY ALONE AGAINST ARAB WORLD
```

Fig.17. Abstracts indicating event/interaction chronology for June 1967 for the Actor countries USA, USR, UAR, and ISR. Decode values for SOURCE, EVENT, and ARENA shown.

Plots of Frequency Distributions

It is often of interest to be able to visually or graphically show frequency distributions of event/interactions by specific countries over a limited time period. As an example, plots were made of acts by the countries USA, USR, UAR, and ISR for the 30-day period, June 1967. The required data file was constructed by GIS using appropriate search logic of the Analytic file. The total number of acts found was 319; it might be of interest to record weekly totals:

	FIRST WEEK	SECOND WK	THIRD WK	FOURTH WK
USA	29	19	23	30
ISR	15	34	9	19
UAR	29	16	6	18
USR	17	13	21	21

Table 2. Weekly totals for USA, ISR, UAR, and USR as ACTORS during the month of June 1967. Total acts (event/interactions) for month is 319.

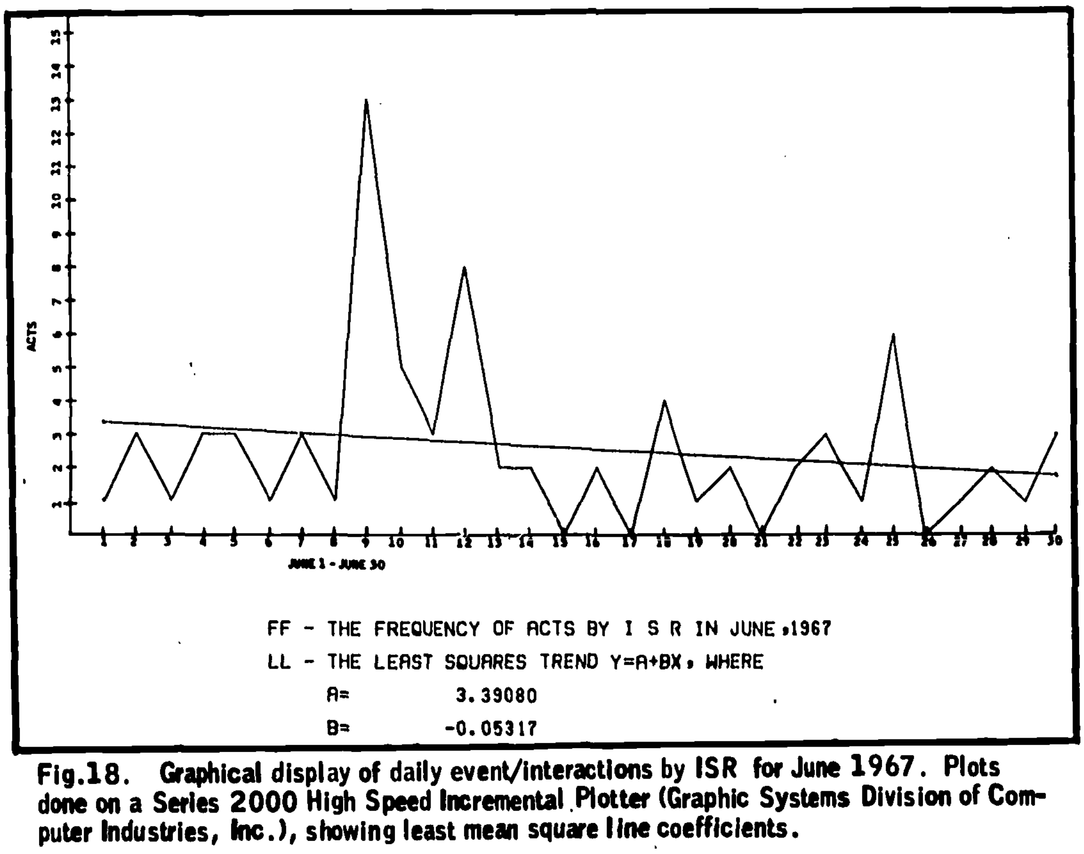


Fig.18. Graphical display of daily event/interactions by ISR for June 1967. Plots done on a Series 2000 High Speed Incremental Plotter (Graphic Systems Division of Computer Industries, Inc.), showing least mean square line coefficients.

The program written to perform the plotting operations and to calculate the least mean square constant and line coefficients was in FORTRAN. The main program performs the plot dimensioning and reads the data set containing the frequency information; a subroutine calculates the constant and x-coefficient for the least mean squares line, $y = a + bx$. The calculated values of a and b are returned to the main program. The graphical output for this combined data retrieval and presentation is shown in Fig.18.

Conclusions

The practicality of retrieving and presenting complex interactions in the area of political science has been demonstrated with the use of several data management and statistical programming packages. Extensions to more sophisticated analyses is possible and may be readily implemented.

Acknowledgements

The work described was funded by Dr. Charles McClelland, Professor of International Relations, University of Southern California, to whom the author is grateful for his kind help and consideration. The statistical processing programs were done by Mr. Verne Tice and Mrs. Tsuey-ing Chang, both of the University Computing Center. Mrs. Chang also deserves credit for the digital plots.

The IBM Generalized Information System (GIS) is a data management program product of the IBM Corporation. The Biomedical Computer Programs, developed at the UCLA Health Sciences Computing Facility, and the Statistical Package for the Social Sciences (SPSS), developed at Stanford University, were used for the statistical processing portions of this work. The plots were drawn using plot routines and a plotter made by the Graphic Systems Division of University Computing Company.

References

1. "World Event/Interaction Survey Handbook and Codebook", Technical Report No.1, compiled and revised by B. Fitzsimmons, G. Hoggard, C. McClelland, W. Martin, and R. Young. Dept. of International Relations, Univ. of Southern California, January 1969. 35p.

MORPHOLOGICAL ANALYSIS WITH PARAMETERS,
INFORMATION CODING SCHEME

SHELDON ISAACSON
MANAGER, CENTER FOR INFORMATION
AND ADVANCED TECHNOLOGY
KOOR INDUSTRIES LTD.
PLANNING & DEVELOPMENT DIVISION

SYNOPSIS

This paper suggests a documentation indexing system based on Morphological Analysis coupled with selected Parameters (MAP). This indexing, based on functions, boundary conditions, and constraints, permits highly structured retrieval. This approach uniquely classifies documents into discrete functional paths - permitting retrieval independent of discipline.

Introduction

Today's rapid technological changes are producing new information and new combinations of information crosslinking all scientific disciplines. This information is highly structured. Most existing indexing and coding systems exhibit low retrieval efficiency with highly structured information. Perhaps one exception is the chemical field where the chemical formulas serve to structure the information quite well.

For many years numerous renowned information scientists have suggested many approaches to this problem of structuring information, such as colon classification, faceted systems, etc. Today, more than ever, these suggestions are relevant. New applications for technical information, especially in technical forecasting, advanced design and analysis and economic projections foreshadow a tremendous upsurge for structured technical information, coupled with constraints, conditions and economics.

Economic constraints and cost-benefit analysis are having a tremendous impact on selected technical situations. No longer is the first solution the most important. It is the best solution that counts. It is most important to have a coding system that extracts and presents this essential data from our reports. The pros and cons of such a system are examined in this report.

Past Experience

As a result of my practical experience with technical information center users at various government, university and industry levels, my views on the effectiveness of present-day coding systems have crystallized. For example, some of the general types of information sought were:

1. What is the state of the art of Function X ?
2. What devices are available for performing Function Y ?
3. What are the specifications, operating conditions and cost for System Z ?

The search for information - some highly structured - was usually carried out by professional information specialists using existing techniques backed by computer support.
Answers were supplied in the conventional manner; once in a while, however, a cross-discipline solution or an alternative path would be suggested which led to some unique solutions based on the function performed.

Functional Approach

Most information requests centered about specific functions within a given discipline such as: detection, amplification, joining, mixing, analysing, etc. Constraints regarding the function, such as cost per function, operational and environmental criteria were desired but extremely hard to retrieve. Figure 1 shows a typical matrix concerning performance of a function. In referring to function, constraint and boundary conditions, we mean

> Function - what is going to be performed ?
> What will be the cost and what are the
> limits of operation and environment ?

If available information could be coded in this manner, its usefulness as well as its accessibility could be increased. Let us analyse some of today's coding systems.

The shortcomings of most of today's indexing and classification systems lie in their inability to provide functional, highly descriptive data to the users. The purpose of the classification system is to group similar items together so that a suitable coding system may be devised for referencing these items.
In the UDC system, for example, automation equipment would be classed under a variety of headings:

620.1	Material Testing
621. 1/9	Machining Details
621.3	Electrical Engineering
621.5	Pneumatic Energy
621.6	Fluid Distribution
621.8	Power Transmission, material handling, mechanical fixing, attachment, lubrication
621.9	Tools, machine tools, machinery

It has been said[1] that each of these headings divide into 50 - 250 subheadings. UDC is too general to be applied to automation equipment. Even with the sophisticated SDI Selective Dissemination of Information system, using descriptors, there are shortcomings in that all fields have to be coded to pick up the same function in any one discipline.

Discipline Crossover

Another factor to be considered is the impact of discipline crossover. In the beginning the sciences were broken down into disciplines for each field of study. Today these disciplines interact with each other in various combinations, forming new disciplines, such as bioengineering, magneto-hydro-dynamics, MHD, etc. This implies changes of search strategy to find complete information in these emerging disciplines.

Another aspect of classification includes scale or level. We have in mind the following levels - complete systems, subsystems, components, parts or elements, materials, substances, atomic or subatomic levels.

1) A Classification of Standard Devices for Mechanization/Automation of Production Processes - W.T. Moodie, R.J. Sury
THE PRODUCTION ENGINEER, June 1969

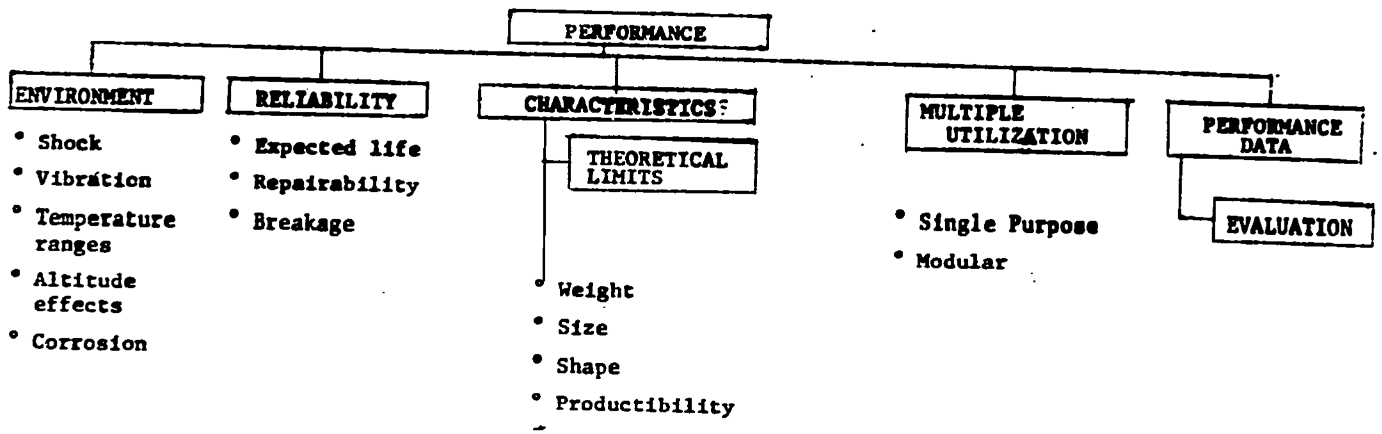

PERFORMANCE
ENVIRONMENT
Shock
Vibration
Temperature ranges
Altitude effects
Corrosion
RELIABILITY
Expected life
Repairability
Breakage
CHARACTERISTICS
THEORETICAL LIMITS
Weight
Size
Shape
Productibility
MULTIPLE UTILIZATION
Single Purpose
Modular
PERFORMANCE DATA
EVALUATION

Figure 1

Communication of Technology

If one asks oneself why we publish so much technical information, one may answer it is to communicate scientific results so that we may advance society as a whole. However, in order to accomplish this, we need to retrieve the data, interpret and understand it and apply it effectively. This new knowledge should be made universally available, regardless of which discipline developed it. One method of achieving this goal is to employ functional coding systems with discipline levels and boundary conditions identified, utilizing morphological methods for deep indexing of our technology.

Morphological Method

The term Morphological Analysis was coined by Dr. Fritz Zwicky. He suggested that, if one studies the features of any object (material or system), its features can be described by a hypothetical Morphological Box. This box would be composed of pigeonholes containing each feature. Zwicky pointed out that this ordered analysis permitted full characterization of known objects as well as characteristics of objects yet to be invented. From this set of characteristics it is possible by permutation and combinations to work out a complete system description. It is also my thesis that this indexing system, coupled with boundary conditions, can fully and accurately index our technical information.

Another aspect of the Morphological Method is that certain paths (solutions) can be eliminated from retrieval by imposing laws of physics and boundary conditions. Further refinements and selectivity can readily be achieved,

If we take a problem that has many potential solutions, we usually analyse the problem by breaking it down into its basic functions and generate alternatives with parameters. Then, in finding the solution, we limit the conditions and select the answers based on optimized parameters.

In the course of establishing the solution "matrix" we can arrive at all potential solutions to a problem. This matrix becomes an indexing tool in that it pinpoints discrete solution paths. The matrix can be used to index technical articles describing a particular solution to the problem. Figure 2 shows the path from problem to solution via the functional parametric route:

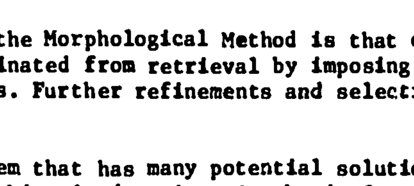


Figure 2

In order to illustrate the operation of the Morphological Analysis Parameter Information Coding scheme, we should examine a few typical systems. Zwicky's method concerns the totality of all the solutions of a given problem. For example, if we were to examine possible propulsion power plants, and if only one specific device, method or system is asked for, this method immediately generalizes the inquiry to all possible devices, methods or systems which provide the answer to a more generalized request. We can also start out with the functions and parameters to establish the matrix. Each parameter Pi will possess a member of Ki independent irreducible values $p^1, p^2 \ldots p_i$. If one element is encircled in each functional matrix and all the circles connected, each resulting chain (isomorph) represents one possible solution of the original problem - this scheme of matrices can be used to construct an N-dimensional morphological box. The morphological method is an orderly way of looking at things and provides a systematic perspective over all possible solutions of a given problem.

Figure 3 illustrates Chemical Propulsion Power Plants:

$[p_1^1, p_1^2]$ = intrinsic or extrinsic chemically active mass

$[p_2^1, p_2^2]$ = internal or external thrust generation

$[p_3^1, p_3^2, p_3^3]$ = intrinsic, extrinsic, and zero thrust augmentation

$[p_4^1, p_4^2]$ = internal, external thrust augmentation

$[p_5^1, p_5^2]$ = positive and negative jets

$[p_6^1, p_6^2, p_6^3, p_6^4]$ = nature of the conversion of the chemical energy into mechanical energy

$[p_7^1, p_7^2, p_7^3, p_7^4]$ = vacuum, air, water, earth

$[p_8^1, p_8^2, p_8^3, p_8^4]$ = translatory, rotatory, oscillatory, and no motion

$[p_9^1, p_9^2, p_9^3]$ = gaseous, liquid, solid state of propellant

$[p_{10}^1, p_{10}^2]$ = continuous and intermittent operation

$[p_{11}^1, p_{11}^2]$ = self-igniting and not self-igniting propellants

Zwicky remarks that " this, if no internal contradictions were present, would make possible

$$\prod_{i=1}^{n} k_i = 2 \times 2 \times 3 \times 2 \times 2 \times 4 \times 4 \times 4 \times 3 \times 2 \times 2 = 36,864$$

Figure 3

If the totality of the paths was considered and no contradiction was present, a possibility of 36,864 propulsion engines could be constructed. However, internal restrictions and physically impossible combinations reduce the number to 25,344 propulsion engines of which only 30 types exist today.

This system describes these engines - both those that exist now and those that can exist. In addition to the basic operational parameters cited, the specifications and environment factors can be considered, as shown in Figure 4. Additionally, the cost factors can be included to arrive at the best economic solution.

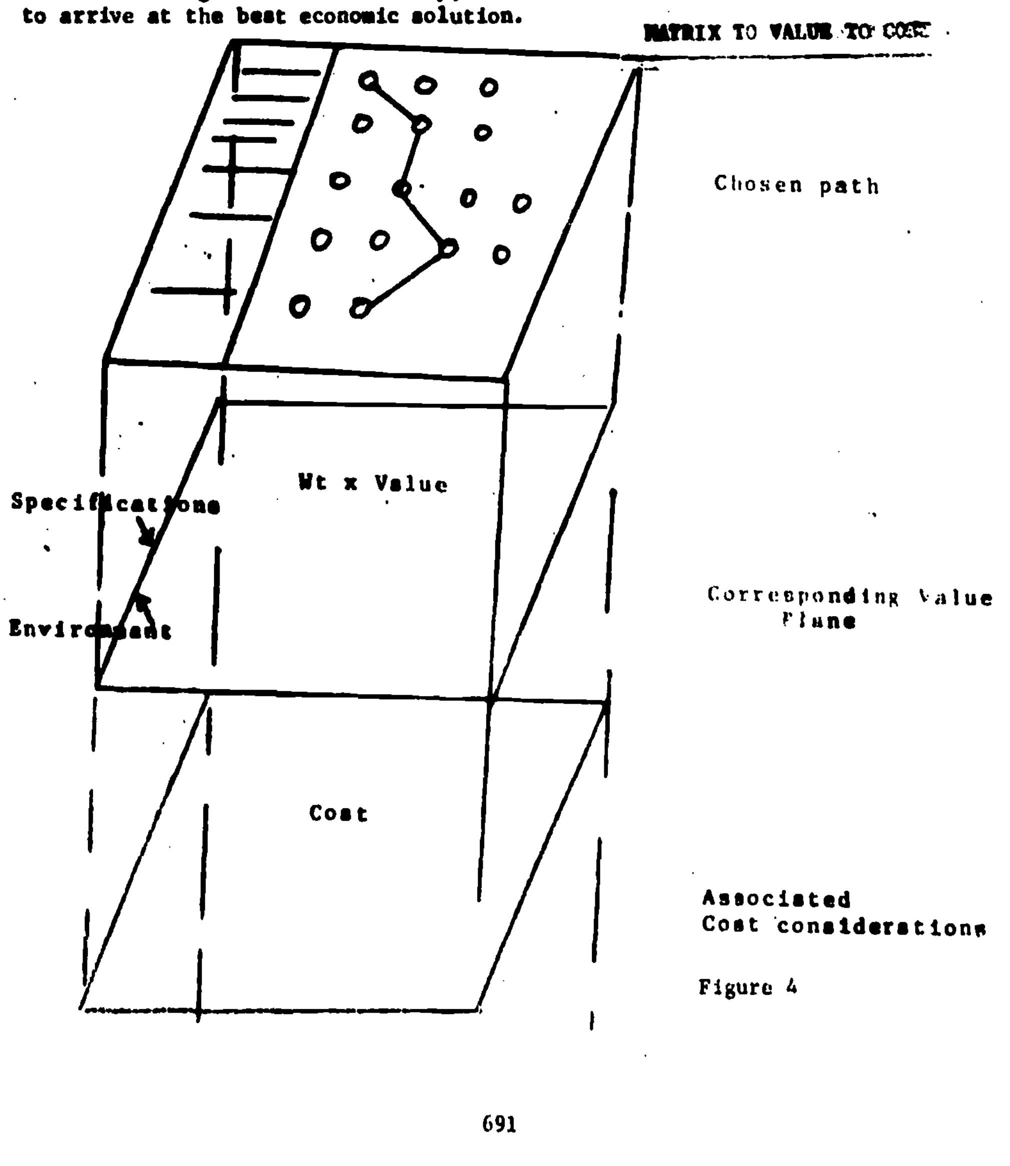


Figure 4

A simplified version of the morphological analysis dealing with the electric clothes dryer was presented by Alger and Hayes, as shown in Figure 5 below:

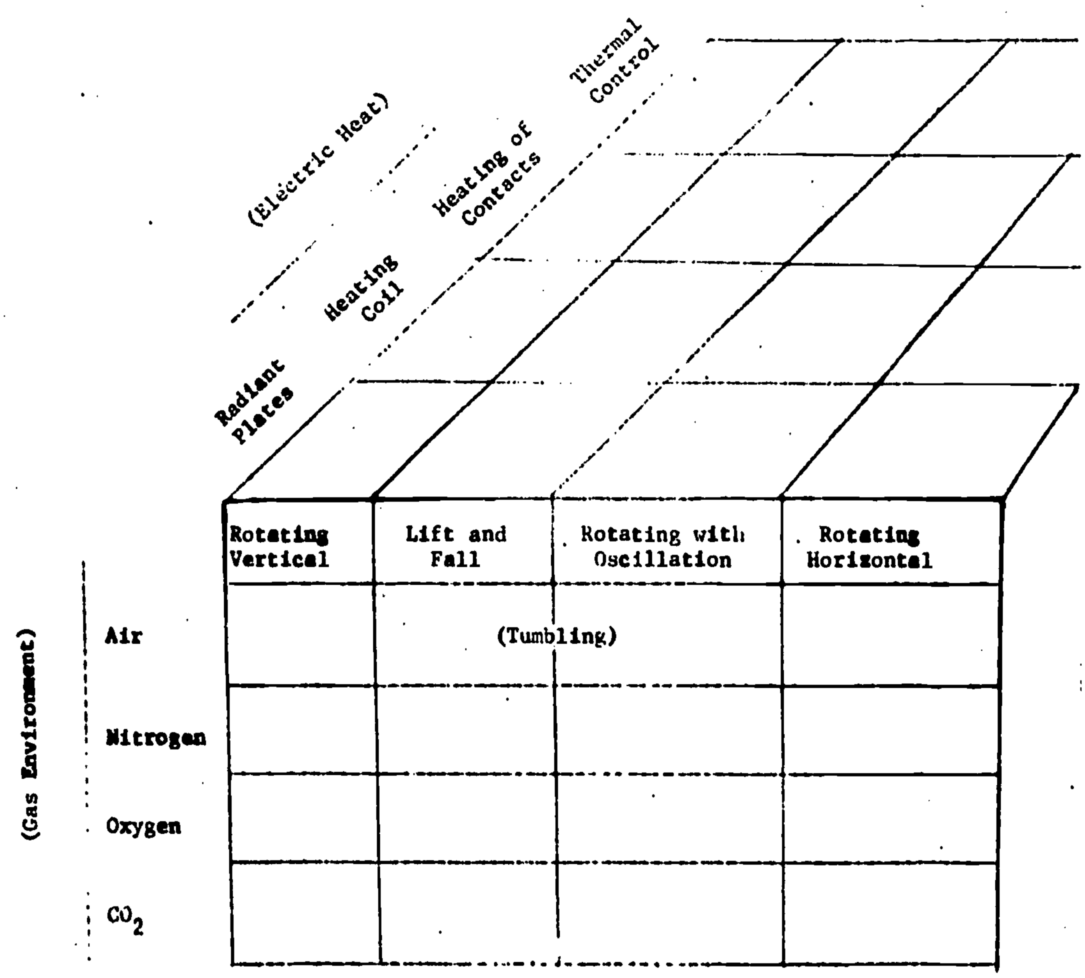


Figure 5

A very detailed application of the Morphology Method was made on the OSCAT[2)] project dealing with the best processing route from an ironbearing feedstock (ore) to a shaped pure iron product.
The metallurgical operation of extracting metal from ore involves three functions:

(1) <u>Separation</u> - extracting and purifying the metal values from the gengue
(2) <u>Reduction</u> - reducing the metal compound to metal
(3) <u>Shaping</u> - forming the metal into useful shapes

One set of conditions is wet or dry processes:

Wet - iron is always liquid or solid
Dry - Vaporized as a volatile compound

(Only wet processes were considered).

A comprehensive list of separation methods is shown in Table 1.
Figure 6 shows the complexity of the problem and the network of alternatives.

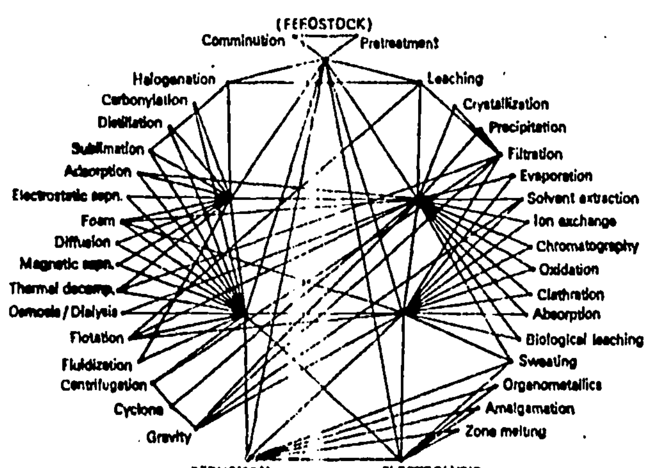


Figure 6.

2) A.V. Bridgewater M.Sc. Thesis, University of Astor in Birmingham, 1966.

(Absorption		G-L
Adsorption		G-S, L-S
Amalgamation		L-S
Biological leaching		L-S
(Carbonydation		G-L, G-S)
Centrifugation	M	G-L, G-S, L-L, L-S
Chromotography		G-G, L-L
Clathration		L-S
(Condensation		G-L, G-S)
Crystallization		L-S
Cyclone	M	G-L, G-S, L-L, S-S
Dialysis		L-L
(Diffusion		G-G)
Distillation		G-L
Electrodialysis		L-S
Electrolysis aqueous		L-S
Electrolysis fused		L-L, L-S
Electrophoroais		L-S
Electrostatic separation	M	G-L, G-S, L-S, S-S
Evaporation		G-L
Filtration	M	G-S, L-S
Flotation		L-S
Fluidization	M	G-S, L-S
Foam		G-L
Gravity	M	G-L, G-S, L-L, S-S
(Halogenation (non-aqueous)		G-G, G-L, G-S, S-S)
(Ion beam		G-G)
Ion exchange		L-L, L-S
Ion excluaion		L-S
Leaching		L-S
Magnetic separation	M	S-S
Organometallic		G-L, L-L, L-S
Oamoais		L-S
Oxidation		G-G, G-L, G-S, L-L, L-S, S-S
Precipitation chemical		G-S, L
Precipitation electrodeposition		L-S
Precipitation electrostatic		G-S, L-S
Production (chemical)		G-G, G-L, G-S, L-L, L-S, S-S
Solvent extraction		L-L
(Sublimation		G-S)
Sweating		L-S
Thermal decompoaition		G-S, L-S, S-S
(Volatilization		G-G, G-L, G-S)
Zone melting		L-S

Some processes are inherent aids (M) to other processes. Thus, the first stage reduces down to 27 possible separation methods. The complete treatment was broken down into 6 stages - complete permutations indicated:

27	Routes	of	Stage 1
729	"	"	Stage 2
19,683	"	"	Stage 3
531,441	"	"	Stage 4
14,384,907	"	"	Stage 5
387,420,489	"	"	Stage 6

Boundary -limitation: 1. Compatibility
2. Economics

What are the implications? First of all, over 387 million classes of papers could be written about ore reduction to final form, each describing some particular route and each presenting an index point. If the boundary conditions are defined, wet or dry process, starting state of metal, etc., this would provide a reduction of routes. Furthermore, if costs-benefits are given per route, a tremendous fallout occurs, reducing the possible routes still further. Figure 7 shows the convergence of routes as a function of applying boundary conditions.

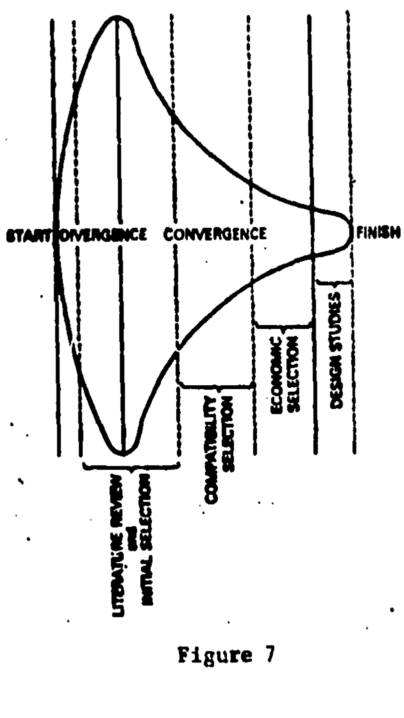


Figure 7

Each of these applications cite exact indexing points for complete retrieval based on functional needs. The system provides for indexing of reports based on theoretical system or systems yet to be invented. Precise retrieval is possible, using this technique.

Coding System

One system proposed for coding the matrix is by functional parameters and their variables.

PROPULSION (CHEMICAL)

$P_{11}21$ 21112 1111

P_{11} indicating 11 parameters

and 21 ... $2 = K_2$ 2nd variable of parameter 1

$1 = K_1$ 1st " " " 2

The position of the number in the chain indicates the parameter. This code can easily be handled by existing computer techniques.

Boundary Conditions

In addition to coding the functions and subfunctions, and their parameter variations, it is necessary to code the boundary conditions. Some suggested factors include:

A. Environment	1. Temperature 2. Humidity 3. Pressure
B. Dynamic Properties	1. Load 2. Strain 3. Stress 4. Strength
C. Mechanical & Physical Properties	1. Density 2. Weight 3. Elasticity 4. Size
D. Theoretical Concepts	1. Fatigue 2. Safety Factor
E. Economic	1. Cost Factors 2. Initial Cost 3. Operating Cost

These factors add a tremendous depth of value to the indexed technical report. It allows one to determine the limits of a system, the cost per function and to compare systems favorably with one another, independent of discipline. This boundary condition coding also provides more specifivity to the retrieval. In other words, given a set of boundary conditions, one could ask for all solutions with emphasis on lowest cost.

Complete Coding

In addition to MAP codes, there would be the boundary conditions codes for each isomorph. These could be coded simply by letter, number and range such as:

A - 1 - 50° C
    2 - 50 %
    3 - 100 PSI
B - 1 - 100,000 #
    2 - 150,000
    3 - 160,000
    4 -   etc.

Constraints on Authors

One of the negative factors of structural indexing is the constraint placed on the authors in determining all these factors. Yet, on the other hand, this deep indexing would force the the author to convey more information about this subject, and would allow users to understand and analyse the report more fully.

Rewards

If all documentation were coded in the same way, the users could readily determine the effect of changes (isomorphs) and determine the results, in effect, using technical information as a system model.

The overall benefits would result in better communication of technology, particulary across discipline boundaries. This would be especially true at the device level where many devices perform the same basic functions, some better than others, for example, the machining of metals can now be done:

mechanically
acoustically
chemically
thermally
optically   etc.

The approach taken depends on the operating and boundary conditions.

Suggested Development of MAPICS

As mentioned at the beginning of this paper, MAPICS is only one approach to structural indexing. However, for this approach as well as the others, certain development measures must be taken:

1. A functional thesaurus must be developed.
2. The boundary condition factors must be tabulated.
3. The system must be widely tested.

Potential Benefits

If the system in its present or modified form proves useful for structural indexing, then certain benefits would be possible, such as:

1. Precise in-depth retrieval
2. Functional retrieval independent of discipline
3. Select retrieval based on boundary condition limits
4. The ability to see "gaps in technology"
5. The ability to cross-fertilize disciplines by "gap filling" analogs
6. The ability to organize the development of a "function"
7. The ability to "predict" or "forecast" new technology based on possible next steps
8. The ability to find wider applications for developed technology in different disciplines
9. The aid to authors in preparing their papers for detailed coverage
10. The ability to "see" small change effects on a total system

THE NATIONAL SERIALS PILOT PROJECT

Tillie Krieger

SYNOPSIS

The Serials Pilot Project attempted to work out the details in setting up a serials data base. Many of the problems encountered resulted from variations in cataloging practice and rules. Some of the solutions are suggested, together with a discussion of some of the basic problems which have yet to be resolved.

Serial literature is one of the most important parts of a library's collection and is indispensable to research in all areas of knowledge. Control of this type of publication has long been of concern to libraries and to scholarly and research communities.

It is difficult to control because its data elements are mobile, often independent of each other and require constant up-dating. A serial is in a fluid state; it changes to meet the needs of its audience, the opinions of its editors and the desires of its publishers. Each issue must be accounted for and estimates of the number of serial titles currently being published range upward from 250,000. Titles are often difficult to identify, describe and locate.

Background of the Project

The National Serials Pilot Project was undertaken to determine the feasibility of establishing a National Serials Data Program. The requirement for this system had been established by the Information Dynamics Corporation study in 1965.

The three national libraries, i.e. Library of Congress (LC), National Agriculture Library (NAL), and National Library of Medicine (NLM), with the cooperation of the National Science Foundation, the Council on Library Resources and the Joint Committee on the Union List of Serials joined to implement the first phase of the National Serials Data Program.

Work on Phase I of the program was conducted by LC and resulted in the MARC format for serials. Work on Phase II was to be more detailed and procedural than that done in Phase I, which was largely exploratory. Each of the national libraries had its own definition of a serial and their practices of handling serials differed. A decision was made to continue with Phase II outside the confines of a particular library in order to be free of the policies and procedures of a single institution.

Initial funds were furnished by NAL. The Association of Research Libraries agreed to act as Executive Agent for the three libraries and the National Libraries Task Force on Automation and Other Cooperative Services provided policy guidance. An advisory committee of librarians from across the country was appointed to advise the National Serials Pilot Project.

Mr. Donald W. Johnson was engaged as the director in September 1969 and Phase II of the National Serials Data Program began.

Objectives and Goals

The objectives stated were:

1. To create a machine-readable file containing bibliographic data on live serials in the fields of science and technology.

2. To produce a number of preliminary listings: a union list, a list of titles held by the national libraries and a list of titles not held by any of them.

3. To produce written reports detailing the findings of the project: problems encountered, solutions attempted with the results, and recommendations for future activities.

4. To develop files, techniques and procedures which would demonstrate the feasibility of a national serials program and which could then form the base for such a system.

The national libraries envisioned the possibility of using compatible data bases for a number of purposes -- determining the extent of their holdings, identifying gaps, locating items not held by one of them as well as showing unnecessary duplication. It could act as a basis for allocating resources and responsibilities among them.

Data could be obtained to assist in planning for effective utilization of their serial collections.

On a broader level it could increase the availability of the world's scientific and technical serials both bibliographically and physically. It could aid in the development of more effective and efficient library networks and information systems. It might simplify and expedite the processing of the literature and it would facilitate the planning, evaluation, and direction of programs and projects involving all of serial literature.

Among the users of the Program would be the national libraries, all other interested libraries, individual users, publishers and distributors of serials, and abstracting and indexing services. Such a central store would allow the sharing of the results of intellectual activity and minimize costs incurred in such a system.

The Base and Format Decisions are Agreed Upon

The Library of Congress agreed to re-write their MARC internal processing programs in COBOL F, to provide source decks and also to provide temporary disc storage for Project data. The Project utilized remote-access typewriter terminals (ATS) which were linked by telephone over ordinary commercial lines to the 360/40 at LC in an on-line mode.

Each week the input data were to be read onto a magnetic tape which was then to be processed on the NLM 360/50 in a one-batch-per-week mode. For practical purposes then, the system was batch mode.

The Union List of Scientific Serials in Canadian Libraries, third edition, was selected as the source or base file. It was on tape and available to the Project. It was a working system which had gone through several editions. The form of entry was established according to the cataloging of a single library, and thus, had a degree of uniformity.

An entry, as defined by the Cataloging Code, is the heading under which a record of a bibliographic entity is represented in a catalog or list, generally the author or title of the work.

The Canadian list had to be reformatted to the ATS output format and from that stage converted to the MARC serials format. At the time it was believed that the Canadian list could easily be reformatted. This turned out to be a false assumption. The structure of the Canadian file had been devised many years ago and the file characteristics were not known.

The reformatting took seven months to complete and, even then, was not completely satisfactory. The MARC format is highly particularized while the Canadian was less so, e.g. data elements were combined in the Canadian list which were handled separately with MARC. In a reformat program, one must be able to identify data elements in the file being reformatted and equate them with data elements in the file to be created. In this instant there was not a one-to-one correlation between elements in the two formats. It was not possible to break out the elements by computer. Going from the general to the particular was not possible and each record required additional editing. In any event as time went on the reformatted file could not be made accessible to the Project through its terminals because it was on the NLM computer, to which the terminals were not connected.

The Project was the first group to make use of the MARC serials format and since they were working in cooperation with LC, became an effective field test for it.

The major decisions as to format and base were resolved in the fall of 1969. In February of 1970 two librarians were added to the staff, which until then had included only the director and his secretary. It was time to begin developing procedures and testing theories.

Work Begins and Basic Problems Assert Themselves

While the reformatting of the base file was underway, another list of titles which had been reformatted to the ATS output format was made available. This was a list of serials indexed in Index Medicus, a total of 2243 medical titles. The staff compared this printout against the printed volume of the Canadian list. All titles which were found would ultimately be entered into the Master file via work on the Canadian list. Titles not on the list were to be entered on a custom basis.

Approximately 275 titles were thus segregated and the process of bibliographic verification, or searching began. It was necessary to search for and verify the various data elements for content and form. This would lead to each bibliographic entity being described in a unique way which would be identified in the same way on all lists and in all files. Since the 275 titles were specific to medicine it was not surprising to find that LC had records for only about one-sixth of the titles. Once searching in two libraries had begun, problems immediately arose.

The Task Force at this time directed that LC files were to be considered prime authority and that all searching was to originate there. The choice for entry, as well as the form of the entry were to be taken from LC files. Only if the title were not found in those files were other sources to be used. It was hoped that this procedure would result in an authoritative serials data base.

A Basic Problem: Entry

Since LC was considered to be prime authority and knowing that they had the largest collection, it was assumed that the number and degree of inconsistencies might be reduced. Other sources generally followed the examples set forth by LC.

Catalog rules, or the manner in which a bibliographic entity is identified, were formulated for the purpose of making the content of the collection available to readers. It serves its purpose best if done according to an established system.

Various access points leading to material are the author(s), and title(s). These are the entries. When one speaks of an entry there are really two facets to the term and each should be considered separately.

The first deals with the manner in which the entry is represented, which rules to follow so that a consistent approach may be followed and a manner of representation stylized. The second facet concerns itself with the ordering of these entries so that the main, or primary entry may be identified and given special status.

A "main" entry is the fundamental or primary heading under which a record is represented in a catalog. The concept has meaning when one wishes to make a single entry listing, where each item can be represented only once. It has meaning when one wishes to produce a catalog card. The concept loses that meaning in a machine file where a list can be specified in any order from a single input. Yet the rules which govern the basic records from which the machine file is built cannot be ignored.

Throughout the history of cataloging the method of determining the main entry has changed. When a change in the main entry occurs, be it author or title, the work has been re-cataloged. A change in a secondary or added entry might involve up-dating, an easier and less costly procedure than re-cataloging.

Since the Project used the results and decisions of cataloging procedures to enter the material into the machine file, conflicts developed over the need to re-catalog which meant a new record, or the need to up-date. This leads to the subject of successive vs. consolidated entries, about which more will be said later.

The problem regarding the form or representation of the entry has not been consistently followed in each of the libraries. In 1967 the Anglo-American Cataloging Code was issued and shortly thereafter adopted. This involved a new approach and led to a different type of representation. An example:

old rules--	American Library Association. Public Library Association. Armed Forces Librarians Section. Military Library Standards Committee.
but under new rules--	Public Library Association. Military Library Standards Committee.

This is only one of the variations possible when working with the new rules as opposed to the old rules. Yet the catalogs in all libraries show the effects of various rules and rule changes over many years.

The decision to accept LC also meant accepting their policy of "superimposition". This means that if the library has already established a form for an author, personal or corporate, it will continue to use that form. It will continue to use that form even for new works produced by that body. Should the body change its name, which would involve cataloging a new work, then the newly done work would be represented according to the latest rules. All names new to the library's collection would be formulated according to the latest rules while previously established names would continue under older forms.

This results in having at least two sets of rules in a single data base. The only way of reconciling the difference is to develop an authority file in the new serials data base.

An authority file would establish the form used, variations to that form would be entered into the file with a guide to the form selected:

I.S.L.I.C.
see
Israel Society of Special Libraries and Information Centres

An authority file enlarges the possibility of locating the desired material regardless of the approach of the user and does not require that he know the rules being employed at any time.

An authority file built at the same time as the master file of serials would permit the most consistent file, as well as one which can be drawn from as it is being built. It cannot be built prior to the file upon which it is based since there is no way of judging what will be needed. It can be built later, but this would involve looking at records a second time, a more costly procedure. Errors which have been made in the absence of such a file might never be detected.

If there is not time, or money, to build a complete file with its necessary adjuncts now, how will there be time to go back and build what is needed and correct the errors made in the meantime?

The MARC serials format attempts to work with the identifying elements of the particular work. It does not include an authority file: it does not preclude the use of one.

Very early in the life of the Project, Mr. Johnson recognized that an authority file would have to be built. He suggested at that time that each record carry its own authority. Once searching began it became apparent that this would lead to an extremely large file and not be as efficient as one would like.

The Pilot Project was charged with the responsibility of listing material as they found it. They had neither the means, staff, nor the authority to change or up-date what they found. They were to present the problems to the Task Force.

Resolution of Other Problems

The manner in which other data elements were found presented inconsistencies for which the Project developed procedures. These included:

Modified letters (diacritical marks)
Capitalization
Language of Place and Volume designations
Holdings information

Modified Letters

Some of the sources checked contained diacritical marks, others did not. Some sources had a variable approach. The staff did not always recognize when the marks had been eliminated and the file being produced has at least 35 different languages represented. The Project people discussed the variations found and the fact that the Project was employing three machines, each of which had a different character set. They reasoned that the marks originated as sounding devices and questioned the need for them. Checking for each language would have involved much time and cost. They concluded that elimination of all modifications would be the most useful, since it would be consistent.

Capitalization

The Task Force requested that the output be in upper and lower case. The sources of bibliographic data are inconsistent but no rules for capitalization have been developed which the Project could employ. This led to the acceptance and transmission of data as they were found.

Language of Place and Volume Designations

Place and volume designations varied. At times they were in the vernacular, or the language of the material; and at times in English regardless of the language of the piece. It was not always possible to locate the form in each of the many languages The English form was easily determined. The serials were not readily accessible and one could not assume what they carried. The decision was made to enter uniformly place and volume designation in the English form.

Holdings Information

An important objective of the Project was the inclusion of holdings. When holdings were easily available, such as on the Index Medicus list which included detailed holdings for each title, they were included. Checklists were sent to the libraries to indicate what titles were contained in the Project Master File, and such holdings as were provided were entered.

At time the holdings were entered in detail and at times in a code which indicated the fullness of the holdings. By means of the code developed by the Project it would be possible to know if the holdings were unknown, complete or substantially so, incomplete, held for a limited time, or if received but not held.

Entering holdings on a detailed or specific level would require constant up-dating and a continuous inventory on the part of the member libraries. It would not be possible for all of the participating libraries to up-date in a constant manner, nor would it be possible for the system to enter such data. The increase in file space would be prohibitive even if only a few libraries were to be represented.

It might, however, be possible perhaps for the local members of a network to carry holdings in a more detailed fashion while the national data bank listed only the code for a library indicating it held a particular title.

Successive vs. Consolidated Entry Approach

A consolidated approach to the cataloging of serials means that the serial is cataloged under the latest title with the complete history of the serial included in the notes. This approach is also referred to as the "entity" approach.

Successive entry cataloging does not involve re-cataloging each time a serial changes its name. Each distinctive title is cataloged with a note made to indicate the immediate predecessor and successor. More items are cataloged, but fewer and less complex notes are required. Also, fresh cataloging is simpler than re-cataloging as would be the case with a consolidated approach.

The MARC format for serials can handle either approach equally well. Looking forward to the probability that a Standard Serial Number (SSN) would be affixed to each distinctive title, the Project took the successive entry approach. Each title was entered as a separate record with machine links established so that the entire history of a serial could be represented as a bibliographic chain.

Standard Serial Numbering Experiment

The Project was not charged with the responsibility for working out details for an SSN proposal. But a complete serials data system would include an SSN. The format allows a tag for the use of one which will be used when the details have been worked out. The Project used a tag designated as a Local System Number. Building a number based on the ANSI approved SSN, they used the number to experiment with the linkage structure within the format. Links can be made to preceding, succeeding and all associated records.

Up-dating

Due to the temporary nature of the Pilot Project, no attempt was made to organize an up-dating procedure. Any continuing project would have to make provision for this aspect of maintaining the files. As a network comes into being and develops the up-dating procedures would be developed and strengthened. The Project could not undertake this area of development.

Procedures of the Project

The period from September 1969 to November 1970 was largely spent in learning the nature of the materials with which they worked and in developing procedures.

The procedures used to gather data were, in the main, simple. The bibliographers began by using the Canadian list as the base, and choosing a live title, would examine the various records in the files at LC. They would photocopy the catalog card when one was available, or copy the information they found. When necessary, they would seek additional information in other files.

The editing procedures which followed, entailed indicating the tags, indicators, delimiters, subfield codes and filling out the fixed field data required by MARC. All information was entered according to the MARC serials format and the MARC editing manual, with certain additions made by the Project staff.

The ATS operator would then enter the data into storage. The work would be revised and corrected and held. Once a week all of the work done that week would be sent to queue. The data was de-queued and transferred to magnetic tape. The tape would be picked up at LC and taken to the computer at NLM for processing. The additional material would then be added to the Project Master File.

Specialized diagnostic listing were then produced, complete with error messages. These would be gone over in the office and errors corrected. Procedures for correcting the file have been included in the program package.

A programmer joined the staff in September 1970 and the work of completing the programs and providing the documentation went forward rapidly.

From November 1970 until the end of the Project, June 1971, the time was spent in attempting to go into the demonstartion stage. This led to a substantial amount of file building. Two editors and an ATS operator were added to the staff. A second port to the LC computer was made available.

The 275 titles from Index Medicus resulted in a file of over 400 records as associative titles were found and included. As each record was searched all previous and succeeding titles as well as supplements were also searched and added to the file.

It was with these related records that the Project was able to establish and test procedures involving the linkages provided by MARC.

Major concepts were stated in the formation of the Pilot Project. Working with the details has produced refinement in those expectations. The technical feasibility of the initial program has been proved.

Mr. Johnson has just completed his final report, and this report, together with the systems flow-charts, documentation and procedure manuals provide greater detail than were available at the time of this writing.

Solutions to the basic problems must be arrived at before work proceeds. The problem of entry and the consideration of the formation of an authority file might well be solved if a small group of individuals, faced with the immediacy of the problem, and each with a knowledge of serials cataloging and computer technology could be formed to work out the details and make the necessary decisions.

The capability is here, the need is felt, may we now expect the decisions?

THE COMPUTER AND THE M.I.T. LIBRARY

Myer M. Kessler
Assoc. Director, Libraries; Director, Technical Information Program (TIP)
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

SYNOPSIS

The Computer Applications Division of the M.I.T. Library supports a Technical Information Program (TIP) that is engaged in the design and operation of computer facilities to serve the information needs of the teaching and research community on our campus. TIP is a general data and text management subsystem which serves as a tool for search, retrieval, editing, formatting and general manipulation of organized files such as data, bibliographic lists and mixtures of text and numeric information.

In the past three years an information system has been operational at MIT which was originally designed with a considerable experimental component in mind. We were not constrained by the requirement that it be useful immediately.

After the period of experimentation and design, the system as of two and one-half or three years ago, was imbeded in a rather mature computer and communications network on the MIT campus. We were operational entirely on the time share, remote access system known to some of you as the MAC system. It was a central computer with some 250 or 300 consoles distributed largely in offices, laboratories, and dormitories around the MIT campus, although there were several, a dozen or so, in people's homes and two or three in other cities. Largely, it was a computer communications network on the MIT campus. This was a generalized computer facility. That is, it was not designed or intended primarily for information purposes. It was intended for education, for research, for program writing, and computing.

On this system, we constructed the additional function of information access and to begin with we used as our experimental medium what is a rather sizeable data base in the physics literature. I will not describe it to you except to say that it consisted of titles, authors and the usual information needed for access plus some rather unique features.

At any rate, considerable prior experimention by us and by some outside judges convinced us that one could do a respectable job of on-line retrieval and of an interactive mode available to the users. So the computer environment was a time-shared remote access system without a public and private memory sector. The literature was well beyond the critical size; that is, five years of physics literature, of some 40 or 50 journals. The language of interaction was natural language, fairly convenient for us. It did not require any programming experience and people took it very well. The in-put out-put machinery was largely a typewriter console although we had about a dozen or so cathorayscopes in various places.

The MIT community is quite different from many of the customary user communities in that it is a University community, rather similar in size and function to the Weizmann Institute or the Technion but with important differences. MIT, although it is an engineering and science school, is somewhat different, somewhat broader in its educational policy. We do have a school of music, for example. We have philosophy and literature, economics and so on.

Second, MIT is essentially an educational institution and everything is directed at that. Even the research that goes on at MIT has a very large educational component so that our measure of success had to include student participation in the system.

The other important factor is the method of payment. From the beginning we decided that at least for a while the payment scheme would be that of the partial subsidy; the setting of the system, the design, its maintenance and the services of data base will not be pro-rated and charged to the customer. The only thing that the user would pay for would be to pay for incremental costs of the actual time used in his interaction. So it was not free to the user but he did not by any means pay the entire cost of the system.

When this was first presented to the MIT community, we were in a state of great euphoria because it looked as if the world was going to beat a path to our door. Literally hundreds of people used it, including students, but it turned out - it was obvious within three months - that there was a great novelty effect in the thing, and that the user population was beginning to thin out and within six months we settled to a much smaller incidence of users, which gave us some bad moments. But, on the other hand, this small curve is beginning to grow slowly. So the on-line response was rather satisfactory.

However, it became clear within six months and in fact this has been reinforced in the past two years, that in the on-line use of the system what we expected did not happen, and some very surprising things did happen in the user behavior pattern.

We thought that there would be a lot of interactive manipulation of the literature, and we all had plans and expectations for on-line interaction and man-machine interaction. I don't know whether that is a local situation and to what extent it will be repeated in other places, but this did not happen in our case. What happened was that the user very soon discovered that his needs don't change from month to month; that after a period of experimentation he can design a fairly acceptable and satisfactory profile, or method of inquiry and that it doesn't change at all. The user finds himself asking the same questions month after month as new data is loaded. He very soon learned that it doesn't pay to sit at the computer. The facilities are available to leave a standing order, so to speak, in the computer and every month as the new literature is loaded, it is swept through these various profiles and whatever matches the profile is merely deposited in the private memory of the particular user, and he can go away for six months and find the answers in his private file.

Intimate interaction with the literature did not happen and I should say this was universal. It was obvious within six months that people will not sit and converse with the literature, and the mode of operation is tending in the direction of SDI, although to be sure there is no human interaction. The profiles are designed and adjusted and manipulated by the users.

As this went on, it became clear that it is too expensive to maintain long backlogs of literature for retroactive searches because physicists did not use is much, and so we removed the bulk of the literature from the storage, from the direct access, but we announced to our people that any time they want to make an retroactive search, we will load the tapes and I must report to you that in two years we have not had a single request to do that. In chemistry, on the other hand, there is considerable call for retroactive information. But the physicists behave as if there is no history.

Now, other things began to appear. Immediately word came to us from the users: "look here, every month I am getting a response from the computer. I don't want to be receiving these little print-outs. I want to save the results on my own private memory and merge the monthly retrieval results as they come along." And this was an on-line interaction. We did not do it for the customers. We said, very well, we will develop a merge program and you do it yourself. And so the users began to accumulate a private library of their own consisting of the results of the monthly retrievals from the general literature.

The next request went something like: "very well, we now have results of retrieval from the generally available line. However, we go to conferences, we speak to people, we get letters, we have some information sources that come to us from other roads, from other channels. We want to incorporate those information sources into the results of our monthly retrieval or monthly SDI." So we had to develop simplified input techniques, editing techniques, up-dating, so that now the user could not only accumulate the results of his SDI retrieval on the computer but he could also add information, documents or even comments that came to him from other sources and incorporate that into his file.

What you see here is a trend for every user to develop a private library, independent of what ir in the central library.

The next tendency that we noted has great implications, and that was the use of the computer as a communications device rather than as a text manipulating device. In the field of physics, for example, there are at MIT at least a dozen groups working on various aspects of solid state physics, who said to themselves: "let's get together and merge all our accumulated information and get rid of the duplicates and create a sort of grand MIT file on solid state physics." This is a communication device because in a sense they published a bibliography for the small audience of the MIT solid state physicists.

This is the way the on-line experience grew, rather than through intense, intimate interaction with the literature. It grew in the direction of information manipulation, editing, sorting, adding, merging, rather than in an intensive aearch.

Other things happened as time went on. It occurred to people that a text handling program such as we developed for information handling, can be used for other purposes. And so a professor of physics, who also happened to be the associate dean of the graduate school, said to himself: "if I can have a list of physics articles and retrieve what I want, why can't I also use the same program to have a list of graduate students.

details - age, name, department and so on, and use the same program for retrieving that. Requests came to us to generalize our programs, to handle things other than what we would ordinarily call document information because there are many files of information that are structurally or formally very similar. For example, the list of all the buildings at MIT and all their facilities, how many elevators, how many classrooms and so on. Formally that file is not different from the list of all the articles in some field of physics, and this has led to a broader and broader use of the program and more and more requests to expand the facilities. For example, the use of our retrieval program in non-literature search led to requests that we supply some primitive means for numeric analysis, e.g., to get the average of the salaries of professors or the average age of students. And so into our system went some minor arithmetic capacity.

The surprise was in the use of the on-line; it did not develop in the direction that we expected, the intimate search and manipulation of the literature files, but rather in the sorting and arranging and formatting of information, adding new information, merging and so on. This is what saved the system from extinction, because had we insisted that the user must sit and interact with the literature, I have a feeling that would not have happened and the system would not have gained acceptance.

I should add one more expansion of the system, in two directions. Within the university, there are many disciplines that do not have national services, for example, musicology. We have to provide facilities for day to day file creation, editing, and formatting of these smaller enterprises whose needs are nevertheless to be satisfied in a university. We have been asked to do text analysis, such as frequency of words, word associations and so on, rather than retrieval. What I am describing here is a total system, the same set of programs built in such a way that they can be put together to perform various functions. This generalization of functions, their broadening, contributed to the acceptance of our system.

We are using the same text management system in the area of library management, acquisitions, serial control and so on. Here again there was an expansion of utilization because acquisitions in the library, is really 90% what business managers call purchasing; if you develop a system that handles library acquisition of books, the same system can be used to manage the buying of chalk and erasers and pencils and papers.

This is the road we took to broaden the functional utility of the system in order to make it economically feasible.

The system now is operational and it is beginning to look as if its initial goal, which was to provide an information retrieval system is at least economically a by-product of many other functions, that are really supporting the system financially, and we are in a way getting a free ride in the matter of information retrieval.

We are now in the process of getting on to the third generation computer, so I have gone through the agonizing experience of writing programs, developing systems and getting them going, three times. The computer industry and the computer engineers keep telling us that the next generation of the computer will be cheaper, faster, bigger - especially cheaper. It is obvious to everybody except those people who use computers that they are right. For the user of the computer, it takes two years to develop a functioning system on a new computer - the programming, the arrangements, the management aspect, the introduction to the user population. That is a costly enterprise. And although it is clear that in some of the newer computers, given a three-year setting up period, the subsequent life will be more comfortable and cheaper and faster, the fact is that after three years, they will come and say, we now have a new one. We no sooner develop a comfortable environment in a computer and start profiting from its capability, then the industry introduces a new set of computers and we have this initial, costly, painful and annoying work all over again. I think we have come to the state where we must say, no, we will not accept new hardware. We have a set of problems that are facing us that can very well be solved and observed with the old hardware and the problems that I am talking about are educational - I don't want to use the word sociological because it's been used a great deal - it is the human configuration that has to be built into the computer and that can very well be done on the second generation computer. We need not go to the third. We are going to live with this computer for some time and we will resist very strongly any attempt from the computer engineers - and at MIT they are a sizeable lobby - from enticing us to a better, faster, cheaper machine, after this.

We want to stabilize this aspect of the problem because there are many others that have to be looked at.

Another problem that is very serious is the multiplicity of the data. This has been referred to again and again. At MIT, chemistry abstracts, physics abstracts, Euratom are all showering us with data bases. There is a great deal of duplication. Then a group of three or four professors comes and they say, "look here, we are starting a new program in ecology. Everybody is excited about the quality of life and MIT should do something about it. The students are demonstrating, protesting. There is money in ecology." What is the data base for ecology? Chemistry, physics, engineering, economics, politics, aesthetics, everything.

How can I respond to this type of need? What has happened is that this has become a national problem and sure enough there is a group in Raleigh, North Carolina that is going through all these various data bases, and making a tape called ecological abstracts. This simply cannot work. What is happening is that we are constantly repackaging the same data bases. Perhaps we have to develop a technology of review writing, of collections of papers, and present that to the user population rather than the raw data of abstracts and indices.

The other lesson that we have learned is that although MIT is a sizeable scientific community, it is not large enough to support many rather obvious and useful applications. The scaling factor is such that we cannot possibly run an economically self-sufficient application if we apply ourselves to MIT alone. For example, the manipulation of the census tapes which was completed in 1970 is a huge complex of data. Sociologists, economists, etc. are very much interested in it, but at MIT perhaps 50 people want to use it and the cost of processing it is huge and simply cannot be done on a local basis. We must regionalize a great many of our activities.

In the States we find that the choice of the regional group is very sensitive; we are experimenting with conglomerates, such as Harvard, MIT and the Boston group of universities. There is a beginning of that.

ACCESS - AUTOMATION PROGRESS AT THE ARGONNE CODE CENTER*

M. K. Butler, C. Harrison, Jr., and William J. Snow
Argonne National Laboratory, Argonne, Illinois, U.S.A.

SYNOPSIS

The Argonne Code Center serves as a central information agency and depository for computer programs primarily in the areas of nuclear physics, and reactor design, engineering, and operation. ACCESS, an acronym for Argonne Code Center Exchange and Storage System, identifies the information storage and retrieval system designed to automate the Center's operation. Implementation of ACCESS is on the Laboratory's IBM 360/75 computer.

I. The Argonne Code Center

In 1960, the U.S. Atomic Energy Commission, in response to a proposal by the Mathematics and Computation Division of the American Nuclear Society, approved the establishment of a central information agency and computer program library at Argonne National Laboratory. Membership of the Division, consisting mainly of physicists, nuclear engineers, mathematicians, and computer scientists, desired a formal center for the collection and transfer of information on computer programs of common interest. The U.S. Atomic Energy Commission viewed the Center as a means of maximizing its investment in computer program research and development, and at Argonne the facility was established as an extension of the Applications Program Library within the Applied Mathematics Division. Today, while the Center remains within this Argonne division, it is a separately-funded activity of the Reactor Physics Branch of the USAEC Division of Reactor Development and Technology, with a European counterpart at the Euratom Research Center in Ispra, Italy, sponsored by the European Nuclear Energy Agency.

The Center has two products — the computer program abstract and the program "package." The abstract describes a computer program, the problem it is intended to solve, the machine hardware required, and the material contained in the program package available upon request from the Center. The package consists of the material required to effectively transfer the program to another installation, perhaps even to a different computer configuration. It includes the computer source deck, run deck, sample problem, data libraries, auxiliary or adjunct routines, and the traditional documentation, such as the computer program report, user's

*Work performed under the auspices of the U.S. Atomic Energy Commission.

manual, and programmer's manual. A program abstract often describes more than one program package since the library collection includes separate program packages for each machine version of a program. The programs in the library are identified by accession number and classified into problem categories according to the type of problem they were created to solve. The current collection contains 306 programs distributed among 18 problem categories as shown in Table 1. These programs represent a total of 432 program packages maintained and distributed by the Center.

	Classification Category Identification and Definition	Number of Programs
A.	Cross section and resonance-integral calculations	29
B.	Spectrum calculations, generation of group constants, lattice and cell problems	44
C.	Static design studies	43
D.	Depletion, fuel management, cost analysis, and reactor economics	36
E.	Space-independent kinetics	11
F.	Space-time kinetics, coupled neutronics-hydrodynamics-thermodynamics and excursion simulations	14
G.	Radiological safety, hazard and accident analysis	19
H.	Steady-state and transient heat transfer	21
I.	Deformation and stress distribution computations, structural analysis and engineering design studies	27
J.	Gamma heating and shield design problems	13
K.	Total systems analysis	1
L.	Data preparation	7
M.	Data management	7
N.	Subsidiary calculations	5
O.	Experimental data processing	13
P.	General mathematical and computing system routines	12
Q.	Radiation effects	2
Z.	Nuclear data	2
	Total	306

Table 1. Argonne Code Center Library by Problem Classification, ANL-7411 Supplement 5

The activity of the Argonne Code Center can be treated in three parts: information services, computer activities, and cooperative efforts. In carrying out this three-part program, the Center staff is supported by a network of "cooperative installation" representatives at participating USAEC laboratories, interested universities, USAEC-contracting institutions, and industrial organizations. Each cooperating installation designates a representative to serve as liaison with the Center. These representatives relay information to the Center concerning programs or requests originating within their installation and apprise colleagues of the Center's program collection and operating procedures.

Information services provided by the Center include the preparation, editing, and publication of the program abstracts, the answering of technical inquiries on the program collection — its use and computer implementation, the compilation of computer program bibliographies, and the reproduction and dissemination of the program packages. Over 5000 program packages, or portions thereof, have been distributed since the Center was organized. A histogram of the yearly distribution statistics through fiscal 1970 appears as Figure 1.

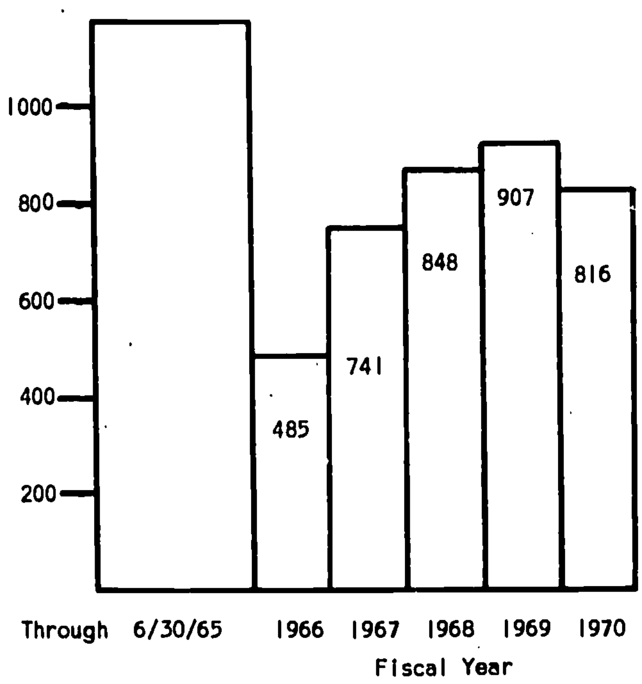


Figure 1. Number of Program Packages Distributed, 1962-70

Computer activities encompass the review and assimilation into the library of the computer-media portion of the program package, the testing, evaluation, and maintenance of the program collection, and automation of the Center's operation through the design and implementation of ACCESS. The collection is composed of program packages offered the Center by authors or installation representatives, programs submitted in compliance with USAEC-contract provisions, those obtained from the ENEA Computer Programme Library through international agreement, and those solicited by the Center to meet outstanding or anticipated requests. The volume of program acquisition activity is indicated in Table 2.

Fiscal Year	Argonne Code Center Programs	ENEA Computer Programme Library Programs	Total
1962	41		41
1963	40		40
1964	48		48
1965	39	14	53
1966	87	16	103
1967	52	43	95
1968	87	34	121
1969	65	34	99
1970	105	10	115
	564	151	715

Table 2. Number of Program Packages Contributed to the Center, 1962-1970

Cooperative efforts cover activities of joint interest undertaken with the participating installation representatives, the Mathematics and Computation Division of the ANS, or the ENEA Library. In the past, these efforts have included publication of a cooperating installation computer facilities handbook[1] and a benchmark problem report,[2] preparation of documentation guidelines[3] and recommended programming practices for computer program exchange, and collaboration on program testing and standardization of library interchange procedures. These, and the other activities of the Argonne Code Center, were described in a paper presented last year at the International Atomic Energy Agency's Symposium on Handling of Nuclear Information.[4]

II. ACCESS - An Overview of the System

In mid-1967, when Argonne National Laboratory put into operation an IBM 360/50-75 system with large on-line storage capacity and terminal editing capabilities, plans were made to automate the assimilation and distribution of the computer-media portion of the program packages and the maintenance of the extensive magnetic tape library. Preliminary specifications for ACCESS were laid out in November of 1967, incorporating statistical and record-keeping functions of the Center along with the library automation. The system as designed provides storage, retrieval, modification, and display facilities for information utilized in the operation of the Center and the performance of Center services. Due to funding limitations implementation of the system was delayed almost two years; programming started in 1969 and progress to date is reported in this paper. The Laboratory computer configuration on which the system is being built is shown in Figure 2.

Any computerized information storage and retrieval system consists basically of two components: the stored information, or data base, and the computer programs written to create, update, manipulate, and retrieve this information. ACCESS contains six interrelated but dissimilar data bases:

1. requests to the Center for computer programs;
2. statistics maintained by the Center on program package transmittal;
3. a directory, or table of contents, of the Center's program package library;
4. descriptive abstracts of the library collection;

ARGONNE NATIONAL LABORATORY
APPLIED MATHEMATICS DIVISION
S/360 CONFIGURATION

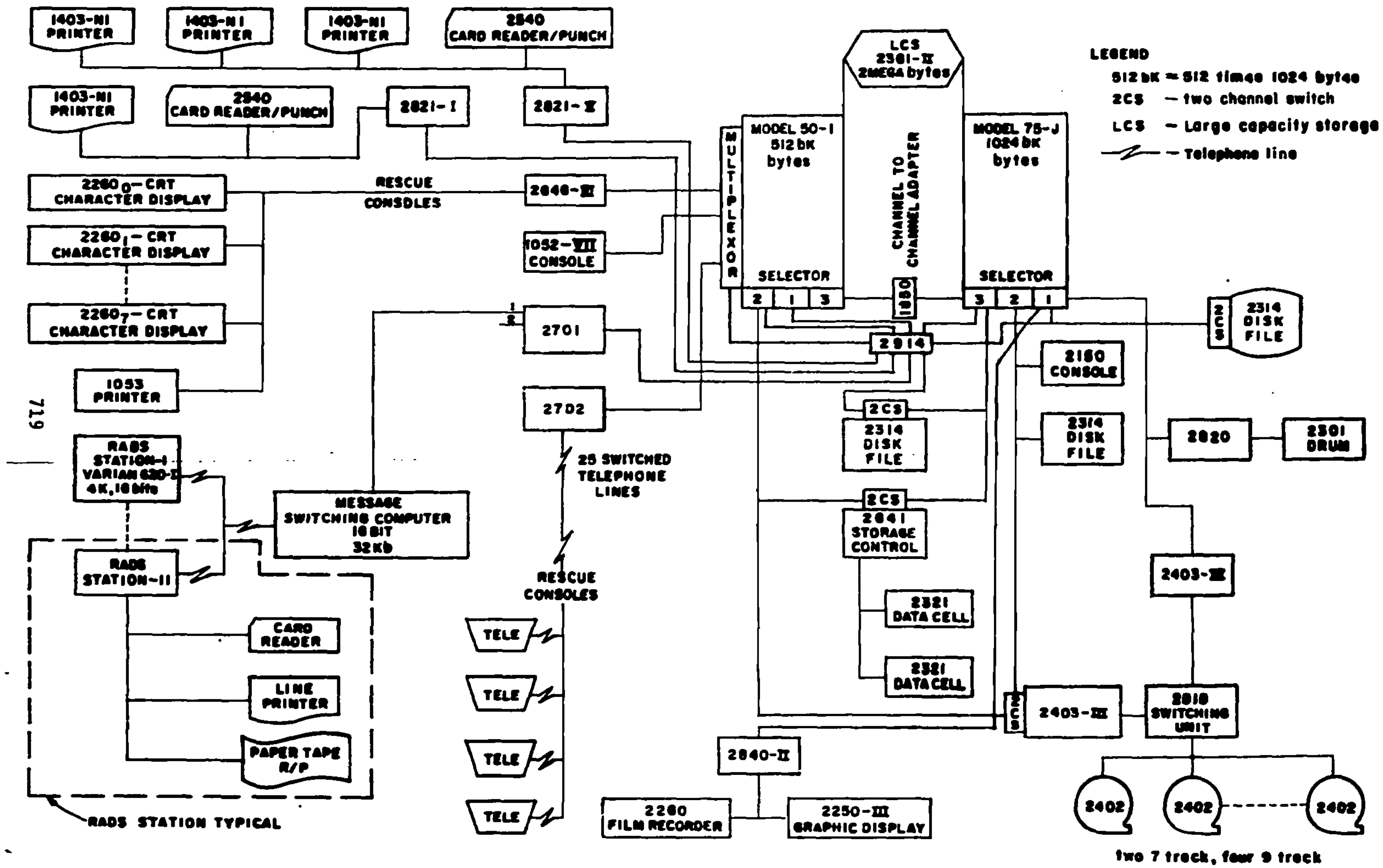


Figure 2

5. library program packages including such items as the source program, object or "run" program, input data for a sample problem execution, auxiliary library data and routines;
6. information on cooperating installations or recipients of library programs, such as hardware facilities, mailing address, library program holdings, and selective dissemination of information (SDI) profile.

Four of these are split into two segments — a "current" segment and a "history" segment to increase the efficiency of the system and to provide an archival facility at reduced cost using off-line rather than on-line devices for information storage.

The system is programmed as a collection of modular programs, each written to perform a specific task utilizing one or more of the data bases. A single program or sequence of programs, or tasks, may be executed at run time. Three types of modules exist in ACCESS. These are identified as the file construction and maintenance modules, the file editing or report-generating moudles, and the special function or task modules. This last type includes many program modules written to automate existing manual or semi-automatic Center procedures, such as the preparation of the semiannual supplements to the compilation of program abstracts published by the Center.[5] A standard ACCESS vocabulary consisting of file-maintenance and report-generating verbs and data-identifier nouns has been defined and is used throughout the system. The commands ADD, MODIFY, DELETE, and PRINT direct the updating and checking of the data bases while the verbs SELECT, ORDER, and EDIT provide the retrieval and editing capabilities. FINIS is used universally to terminate a command string. Data identifiers such as NUM for library accession number, NAM for the program or installation title, CAT for the problem classification or distribution category, ITM for the file entry's item identification, and KEY for SDI keyword or Boolean keyword-combination are invoked with these verbs to update data bases and to prepare publications, reports, and transmittal labels. Few commands are required with the function modules, where programmed conventions and default options are applied; however, the two general terms OVERRIDE and SPECIAL have been adopted to overide programmed conventions when required and to permit substitution of input data for file information on a temporary basis.

A schematic diagram of ACCESS is presented in Figure 3 and a description of the system data bases is given in the following section.

III. The ACCESS Data Bases

As noted earlier, the six data bases while interrelated are quite dissimilar in structure and size. Before proceeding with a description of each of the six individual data bases, a look at the over-all system data base characteristics is in order. By far, the largest data base in terms of storage requirements is the program package collection; it is also one of those most frequently accessed. Four of the six data bases have both current and history segments. In each case, the history segment has been designed for retention on magnetic tape. The two remaining, the request and the statistics files, do not contain archival data. Accompanying each data base is a directory containing the pointers and other control information used by the program modules to access selected entries, or particular items within file entries, to expedite the information retrieval function. The program modules utilize the structure definition provided by the PL/1 programming language to accomplish this. Table 3 summarizes the estimated maximum data base storage requirements for the system and

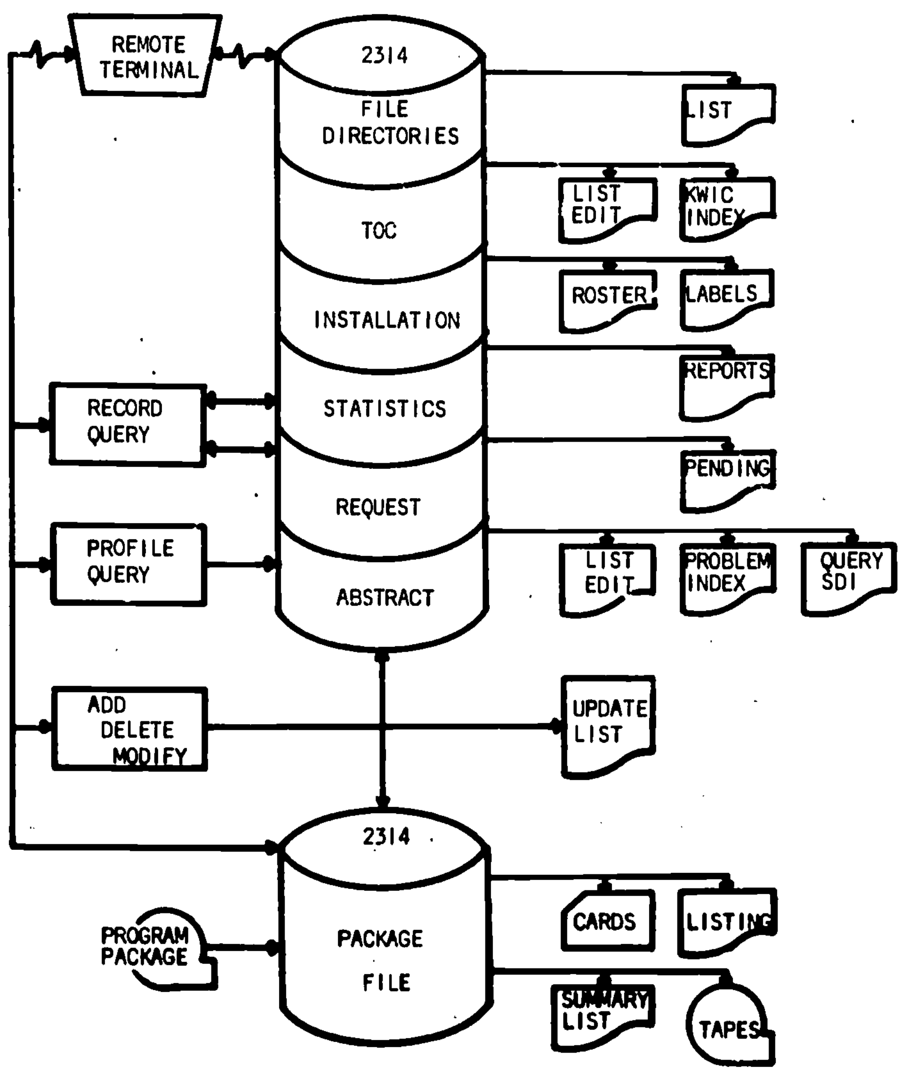


Figure 3. ACCESS Flow Diagram

for each of the individual files, and indicates the size of each established file at this time. To date, no effort has been made to create either the request or statistics data base; this will be undertaken when the package file is established.

DATA BASE		STORAGE REQUIREMENTS				
		Est. Maximum		Implementation		
Identification	Contents	Entries	Bytes	Entries	Bytes	DEVICE
REQ	Unanswered and incomplete requests	2,000	724K	-	-	IBM 2314
STA	Transmittal statistics	4,000	896K	-	-	IBM 2314
TOC*	Directory of program package collection	2,000	240K	432	52K	IBM 2314
ABS*	Program abstracts	1,000	7,200K	306	2203K	IBM 2314
PKG*	Program packages	2,000	300,000K	6	720K	IBM 2314 IBM 2321
CIR*	Coop installations	500	121K	105	30K	IBM 2314
	Other recipients	680	128K	678	127K	
	Library holdings	18,000	144K	5,000	65K	

*History segments of these files reside on magnetic tape.

Table 3: ACCESS Data Base Storage Requirements

The request (REQ) data base was designed to replace a manual "pending" file maintained in the Center office. The Center receives, in addition to the routine requests for copies of material in the library collection, queries on the availability of programs known only by title or by citation, and requests for information on nonlibrary programs. In such cases the staff attempts to locate the appropriate program and to obtain the information, or the program, desired. The ACCESS request file is set up to accommodate as many as 2000 pending program requests and queries at any time. It is a fixed-format file, structured to permit as many as 11 information items for each entry. These information items are:

1. request number,
2. date of entry,
3. cooperating or requesting installation identification,
4. program package identification,
5. package contents symbols,
6. override option codes,
7. special requester's name,
8. address option codes,
9. special address information,
10. special transmittal instructions, and
11. Center comment or miscellaneous information.

Both items 1 and 2 are assigned automatically at the time the request is entered into the file.

The statistics (STA) data base takes up where the request file leaves off. When a request is filled, the request file entry is transferred to the STA file with the date of transmittal appended as an additional information item. Present plans call for entries in this data base over two years old to be purged annually. Archival information will not be stored in this detail. The data base is maintained for two purposes — to provide the facility to check recent transmittal activity to eliminate filling of duplicate requests and to provide summaries of transmittal statistics for Center reports.

Data bases 3 and 4, the table of contents (TOC) and program abstracts (ABS) files, were originally maintained as punched card files to expedite a semiautomatic preparation of the Center's Compilation of Program Abstracts report.[5] The fixed-format structure of these files follows the publication format. The TOC data base permits up to 2000 current entries; each entry consisting of 120 characters. The first 80 characters are the original card format and contain the following information items:

1. a tag indicating BCD or EBCDIC code,
2. a tag indicating restricted or unrestricted distribution,
3. program accession number,
4. program name and description,
5. machine identification,
6. programming language designation,
7. package contents symbols,
8. a tag indicating whether or not transmittal is via magnetic tape, and
9. problem classification category.

Characters 81 through 120 of each TOC entry are allocated for pointers and control information relating to the contents of the corresponding program package entry.

The ABS file has an entry for each program and the current ABS segment allows a maximum of 1000 entries. Abstracts containing fewer than 7200 characters are stored in one section of the file; those longer than 7200 characters are stored in another section. Presently, this data base contains 306 entries and the average entry length is 5760 characters. The 17 information items correspond to the published abstract items as defined in reference 3; these are:

1. program name,
2. machine identification,
3. nature of physical problem solved,
4. method of solution,
5. restrictions on the complexity of the problem,
6. typical running time,
7. unusual features of the program,
8. related and auxiliary programs,
9. status,
10. references,
11. machine requirements,
12. programming language designation,
13. operating system or monitor under which program is executed,
14. other programming or operating information or restrictions,
15. name and establishment of author,
16. material available, and
17. problem classification category and keywords.

The largest data base is the program package file; entries in this file are in one-to-one correspondence with the entries in the package directory, or TOC, file. Program packages are supplied the Code Center by program authors, cooperating installation representatives, and the ENEA Computer Programme Library. The package consists of both computer-media material and the traditional library matter or documentation. ACCESS is concerned only with the computer-media component. This material is received in the form of cards punched with either the IBM 026 code or the IBM 029 code, or in the form of magnetic tape recorded in binary, BCD, EBCDIC, UNIVAC Fieldata, CDC display code, or a mixed binary-BCD mode. These tapes may have been written at 200, 556, 800, or 1600 bpi, in 7- or 9-track recording, with blocked or unblocked format. The system must be able to handle any form acceptable to the IBM 360 hardware. Currently, there is no 1600 bpi tape capability on-site, and when such tapes are received at the Center they must be copied to a lower density off-site. Each package file entry is stored as a contiguous block of records; an entry may consist of as many as five items analogous to the package contents symbol definitions of:

1. source program (S),
2. object program (B),
3. sample problem input (P),
4. data libraries (L), and
5. auxiliary routines (X).

Source program statements are compressed prior to entry by deleting trailing "blank" characters in each statement. Binary programs are stored as byte strings. An entire IBM 2314 disk is dedicated to this data base with overflow directed to the data cell. The system is designed to allocate disk space to the most recent and most frequently-referenced entries.

The cooperating installation and recipient data base (CIR) consists of three sections — the cooperating installation roster, other recipients, and library holdings. This file replaces a manual Rolodex file containing mailing list addressee information, a card index where library transmittals are recorded by requesting installation, and several program notebooks in which the name and address of each recipient of each library program is recorded, together with the date of dispatch and an indication of the material sent. In ACCESS the cooperating installation section is capable of holding 500 entries; each entry consists of the seven information items:

1. installation identification,
2. machine identification,
3. subscription categories,
4. name of representative,
5. his telephone number,
6. his mailing address, and
7. the installation SDI profile.

The other recipients section has provision for 680 entries consisting of the three information items 1, 3, and 6 listed above. Item 3, the subscription categories, identifies the Center mailing lists on which this entry appears. Section 3 of the CIR data base contains an 8-character entry for each program package transmitted. The information stored is:

1. the year of transmittal (fiscal year),
2. identification of recipient, and
3. program accession number.

These six ACCESS data bases are permanently resident on two IBM 2314 disk packs and an IBM 2321 data cell. No redundant information is stored in the system; instead, pointers and control information are utilized by the system's program modules to access the information in the various data bases.

IV. ACCESS Program Modules

This final section describes the ACCESS program modules which have been written in PL/1 or ASSEMBLER language. The modules are kept on the IBM 2321 data cell in both source and object form, and may be modified and executed from a remote terminal. Implementation of ACCESS started with the creation of the TOC and ABS data bases, development of the ACCESS vocabulary, and specification of general file construction and maintenance and file editing and report-generating facilities. Efficient operation of the system requires that:

1. the routine information storage and retrieval functions be performed as rapidly as possible, and that
2. the dissimilarities of the data bases be minimized by adoption of a standard vocabulary and common procedures for system-user communications, thereby reducing the number of operational errors.

ACCESS data bases are structured as fixed-length records on the direct-access storage devices. A directory, accompanying each file, contains the relative track location information for each file entry. When a file is referenced, the directory is brought into core memory and this record address information is directly accessible. Most program modules transfer data base information between core memory and the permanent storage device in full-track units, checking to determine whether a track is in core before searching the permanent storage device, to optimize system performance.

An independent update module services each of the data bases. The most extensive facilities are those provided for the larger program package and abstract files; these utilize the ADD, DELETE, and MODIFY commands. By means of these commands, issued via card or terminal input, the system user may add or delete entries (abstracts), or blocks of entries in the ABS file; he may also modify entries or entry items. No abstract with an accession number identical to an accession number already in the file will be added. When an abstract is added or modified the date of this action is automatically recorded in the date field of the abstract record. The package file update facility allows the user to add or delete entries (package contents) or entry items (components) in the PKG file; he may also modify entry items, or lines and blocks of lines within entry items. Package data may be entered from cards, magnetic tape, or a terminal. When data are entered the corresponding package contents symbols in the TOC file are checked for consistency, and if a discrepancy is found the user is notified.

Updata programs for the smaller data bases do not utilize the ADD or MODIFY commands but rely instead on a "replace" convention. If an entry is supplied, the file contents are checked and if that entry is already present in the file, a MODIFY command is assumed, and the new entry replaces the old. If that entry is not present, an ADD command is assumed, and the entry is added to the file. The system-assigned request number identifies the entry in the REQ and STA files. A "blank" request number implies an ADD command to the request update module and an error to the statistics file maintenance program. The installation identification symbol, or number, identifies the entry in the CIR file, and the program accession number combined with the machine designation is used for the TOC data base.

The DELETE and FINIS verbs are recognized by all file update programs. When FINIS is encountered, the maintenance programs terminate providing a printed listing of the added, modified, and deleted date base information processed. The PRINT command elicits a complete edit of the updated file for all system data bases, except PKG. Data base information prepared on an IBM 026 keypunch is universally acceptable as input, as well as the standard IBM 029 information.

File editing and report-generating facilities have been developed in a similar manner with independent programs servicing each of the data bases. The commands used by these modules are SELECT, ORDER, and EDIT. The SELECT verb extracts, or selects, a particular subset of data from a data base for use as a temporary file. The data-identifier nouns described earlier serve to identify the subsets to be extracted. When the temporary file has been constructed, it is arranged or ordered as prescribed by the ORDER command. The same data-identifier nouns — NAME, NUM, CAT, DAT, ITM, and KEY prescribe the arrangement of the temporary file. When ordered, an EDIT command is issued to output, or edit, the contents of the temporary data base. Titles and page headings may be appended and the number of lines per page specified.

Probably the most elaborate file editing program is the one written for use with the abstract data base. This program permits the system user to prepare the edited program abstracts and the problem classification guide included in the semiannual supplements of the ANL-7411 report. The command string used to produce these edits for the most recent supplement was entered via a card deck prepared as follows:

```
     Card 1  SELECT   NUM=295,329,358,368,411-449
          2  ORDER    NUM
          3  EDIT     TITLE=PROGRAM ABSTRACTS, SUPPLEMENT 5
          4  SELECT   NUM=1-999   ITM=3R
          5  ORDER    CAT  NUM
          6  EDIT     TITLE=VI, PROBLEM CLASSIFICATION GUIDE
          7  EDIT     PAGE(A)=A.   CROSS SECTION AND RESONANCE INTEGRAL CALCULATIONS
          8  EDIT     PAGE(B)=B.   SPECTRUM CALCULATIONS,GENERATION OF GROUP CONSTANTS,
          9  EDIT     PAGE(B)=        LATTICE AND CELL PROBLEMS
         10  EDIT     PAGE(C)=C.   STATIC DESIGN STUDIES
             :        :        :
Next to      :        :        :
last card    EDIT     PAGE(Z)=Z.   NUCLEAR DATA
Last card    FINIS
```

The first card instructs the ABS edit module to select abstracts for programs with accession numbers 295, 329, 358, 368, and 411 through 449 and transfer them from the ABS data base to the temporary file. The second card causes the program to put these abstracts into numerical order by program accession number, and the third card tells the program to print a title page with the legend shown, followed by an edited listing of the temporary file. Completion of the SELECT, ORDER, EXIT sequence terminates action on this temporary data base, and the SELECT on the next card indicates that a new report is to be generated. This second sequence of commands selects item 3 from each ABS entry, replacing the item heading with the program name. These items are then arranged in alphabetical problem classification order, with the items within each category in program accession-number sequence. This second edit produces the problem classification guide. The FINIS command terminates the module's execution.

Analogous input used with the request data base editor can produce a list of pending requests, from a particular installation, over a specified period of time, or for selected program accession numbers. Transmittal activity reports can be generated from the STA file showing fiscal year package distribution, by library program accession number, or by requesting installation identification. Address labels for Code Center mailings can be run off using the CIR edit module, and a telephone directory or roster of cooperating installation representatives is readily constructed from the installation data base using the file edit facility.

In addition to the two general program module types described, there is a third type of program module in ACCESS. This is the function or task module. These are special-purpose programs written to perform a specific task. A good example of this type is the KWIC program module which produces a keyword-in-context index for the ANL-7411 supplements from the TOC data base.

Another task program has been written to process the routine program package distribution requests. Program package information is usually transmitted on magnetic tape because of its size; however, punched cards and listings are alternate transmittal media employed for the smaller programs or upon request. Binary information in the package data base is stored without change and without regard for word length. All other information is stored and maintained in IBM 360 EBCDIC format. Facilities are provided in ACCESS and the COPY program which permit program package material to be distributed in any of three forms. One of these is the format in which the package was submitted to the Center, i.e., the TOC machine designation format. The second is the format compatible with the machine at the cooperating installation, and the third is a special machine format specified with the request. To provide this service a table of machine format characteristics is maintained within the COPY program. Characteristics included are:

1. machine word length (bits/word),
2. binary representation (bits/character, octal or hexadecimal),
3. character sets (card and tape),
4. tape recording specifications (number of tracks, parity, density).

The COPY program first checks the entry in the REQ file to determine if a SPECIAL format has been requested. This option allows the user to specify a record blocking factor and the four machine format characteristics. If a SPECIAL format has not been requested the program will attempt to complete the request using the appropriate installation format. If no installation format data exist the TOC machine designation is used to select the format from the PKG machine characteristics table.

Other task modules have been prepared to perform system utility functions such as disk-to-tape and tape-to-disk transfer and terminal entry. New programs can be readily added to the system as needed due to the modular design of ACCESS.

References:

1. Butler, M. K., Legan, M., Lindsay, E., Ranzini, L.; Argonne Code Center; A Directory of Cooperating Installations; ANL-7497 (1968).

2. Argonne Code Center: Benchmark Problem Book; ANL-7416 (1968).

3. American Nuclear Society; A Code of Good Practices for the Documentation of Digital Computer Programs; ANS-STD.2 (1967).

4. Butler, M. K., Harrison, C., Jr., Snow, W. J.; The Argonne Code Center: A Decade of Computer Program Exchange; Handling of Nuclear Information Proceedings of a Symposium, Vienna, 16-20 February 1970; (1970).

5. Butler, M. K., Henline, P., Legan, M. Ranzini, L., and Snow, W. J.; The Argonne Code Center: Compilation of Program Abstracts; ANL-7411 and Supplements 1-5 (1968-1971).

COMPUTERIZED PRODUCTION OF LIBRARY CATALOGS

Justin M. Kniemeyer
Director of Automation Services
Tarrant County Junior College District, Forth Worth, Texas

SYNOPSIS

This paper discusses the decisions and considerations necessary for the computerized production of library catalogs. General computer requirements for data storage and processing are outlined. Three methods of library catalog production are considered and estimated costs of production are provided for each type.

A desire of librarians for many centuries has been to produce a library catalog that is accurate, complete, easy to update and economical to maintain. Several varieties of catalogs have been proposed and attempted over the years. Each has advantages and disadvantages. This paper assumes the premise that the production and maintenance of library catalogs through the use of modern computer techniques is a valid concept and may be at least a partial solution to some of the problems that plague other library catalogs.

Library catalogs have always been prepared in one of two ways; either in a card format, with all its problems of misfiled cards and difficulties of reproduction, or the bound book format that is out of date as soon as it is printed. Computerized methods of library catalog production eliminate the misfiling problems of the card catalog. The speed of the computer permits new catalogs to be prepared often enough that the out-of-date problems of the book catalog are overcome.

Computers opened new avenues of exploration concerning the production of library catalogs. Computers print extremely fast, amazingly accurate and for long periods of time without having to stop. These qualities alone are enough to arouse the full attention of library administrators operating with less than adequate staffs. The computer, however, is able to assist in other areas concerned with bibliographic control. The computer can store the data relatively inexpensively in machine-readable form, it can arrange the data in any desired predetermined sequence, and it can interfile different groups of data arranged in the same manner at speeds measured in millionths of seconds rather than hours or days. All of these attributes of a computer indicate that its utilization for the preparation of library catalogs is worthy of consideration.

Recording the required data in a machine-readable form, storing the data and processing it for printing need to be the first considerations in utilizing computers for the preparation and production of library catalogs. Existing libraries with large collections face the most difficult decisions concerning how to convert their data to machine-readable form. The most common method is simply to convert the shelf list file to punched cards and then process them into the computer. This method requires a considerable length of time to complete and a staff of trained keypunch operators. For example, in 1967 in Winona, Minnesota, a project called the <u>Winona Colleges Union Catalog Project</u> was undertaken. This project was to produce a union catalog of the holdings of three local college libraries and the city's public library using computerized methods. The total holdings were estimated at 250,000 titles. The method of conversion selected was punched cards. A staff of six full-time key punch operators was employed. Eighteen months and $30,000 later, the raw data had finally been converted.

Other methods of data conversion are being developed and of particular interest to libraries with large collections are the optical character readers. As this technique matures and as automatic format recognition techniques are developed, existing shelf lists can then be converted automatically and quickly.

Various storage mediums are available to store the converted bibliographic data. The three most common and most often considered are: (1) punched cards (2) magnetic tapes and (3) magnetic disks.

Punched cards, as the most often used means of converting data to a machine-readable form, are perhaps the first storage medium considered. In small systems where the total number of cards does not exceed a quarter of a million, it is possible to use punch cards. In larger systems other storage mediums should be considered. Magnetic tapes are perhaps the most common storage medium in use for the storage of large amounts of data. Common tape recording densities are 1,600 bits per inch. An excess of 46 million bits of data can be recorded on one 2,400 foot reel of magnetic tape. Many thousands of titles can be recorded in 46 million bits of data. Magnetic disks are the most recent storage mediums considered for large bulk storage data files. The most obvious advantage of disk storage over other types of storage is the capability of being able to access any individual record on a random rather than on a sequential basis.

In evaluating each of these three types of bulk storage mediums, the type that best fits the needs of library catalog production in a book format is magnetic tape. In utilizing magnetic tape storage, large amounts of data can be stored in a small amount of space. The data is stored in a machine-readable form and the data can easily be updated and processed by the computer. Punched card files, on the other hand, require large amounts of storage space to store small amounts of data. Updating punched card files is an endless and seemingly impossible task and the speed of processing is slow when compared to that of magnetic tape. Magnetic disk storage is advantageous only if the entire file must be on-line continuously. If the system plan is to sort and to print the entire file rather than randomly select and display an individual record, the cost of maintaining magnetic disk files as compared to magnetic tape files cannot be justified.

Storing bibliographic data requires a great deal of concern regarding the record format and the "completeness" of the record. The first decisions to be made are those regarding the identification and specification of exactly what data elements are to comprise the stored record. The storage of the total amount of data contained on cards produced by the Library of Congress in most cases is not feasible, desired or needed. Usable storage space has its practical limits and should be conserved as much as possible. In the majority of educational institutions today, the wealth of information contained on a Library of Congress printed card is not required. An abbreviated form of entry contains all of the necessary data to locate a particular item. Figure 1 is a typical example of an abbreviated entry. By providing an adequate number of indexing points and by assigning subject classifications that adequately reflect the contents of the title, the material will be easily located and the entry will be usable.

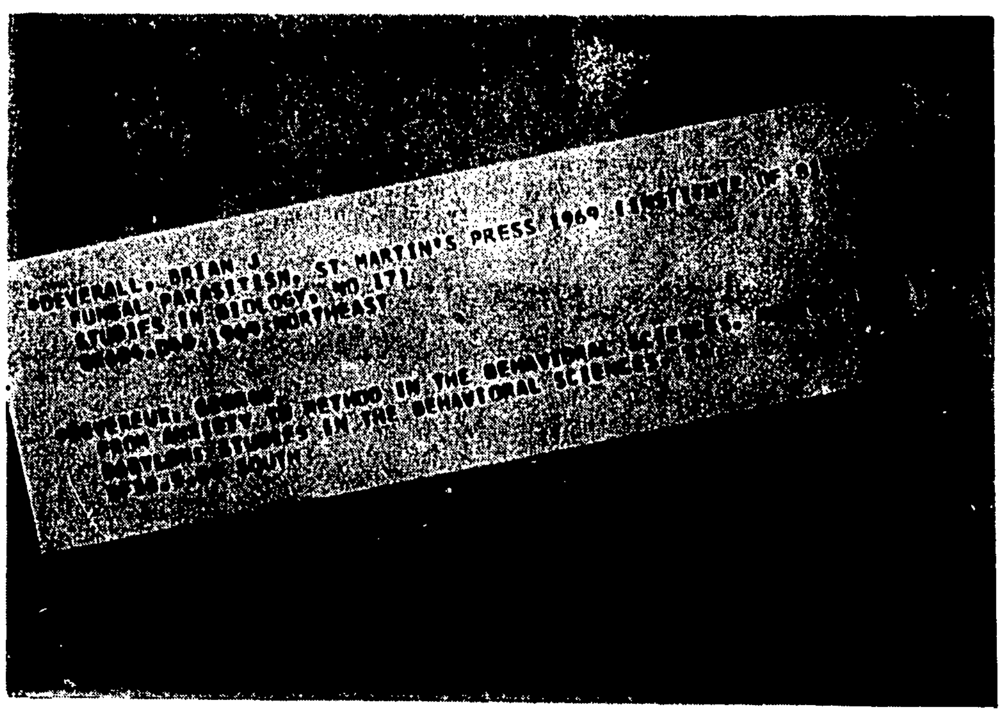


Figure 1

The actual layout of the data record should be performed by a competent systems analyst in order to ensure the most efficient computer processing possible. This processing requires the sorting of data into the proper sequence for printing. At this point a decision must be reached regarding the type of sorting to be performed. The normal computer collating sequence is not desired by many institutions. These libraries must then set a programmer about the task of developing a set of sorting rules to permit filing according to the Anglo-American, American Library Association, or some other set of filing rules. The designing of hese rules depends, almost always, on the input data according to a specific format. If the normal collating sequence of the computer is selected, a specific format for the input of data must also be developed. The use of the computer's normal collating sequence will also permit faster processing times; in that fewer program manipulations are required. What the computer does naturally may be the best decision to make.

The output format desired is the next decision that is required. The most basic portion of this is to determine whether a dictionary or divided type library catalog is wanted. The dictionary approach has much to offer in that records are filed in one alphabetic arrangement and the user has only one source to go to for information. The divided catalog does force the user to choose either an author, title, or subject approach to his problem before using the catalog but the computer is capable of producing the divided catalog more easily that the dictionary catalog. In a dictionary catalog, some clear-cut means of identifying between author, title and subject entries, must be utilized. The normal manner of doing this is with either boldface type or red ink. The computer has neither at its disposal. Dictionary catalogs produced by a computer tend to be difficult for the user to use. For these reasons the best choice is the divided library catalog with separate author, title and subject volumes.

The exact printing format of an entry in each of the three catalogs generally follows the type discussed previously for abbreviated entries. Figure 2 is an illustration of the appearance of a single entry in an author, title and subject catalog.

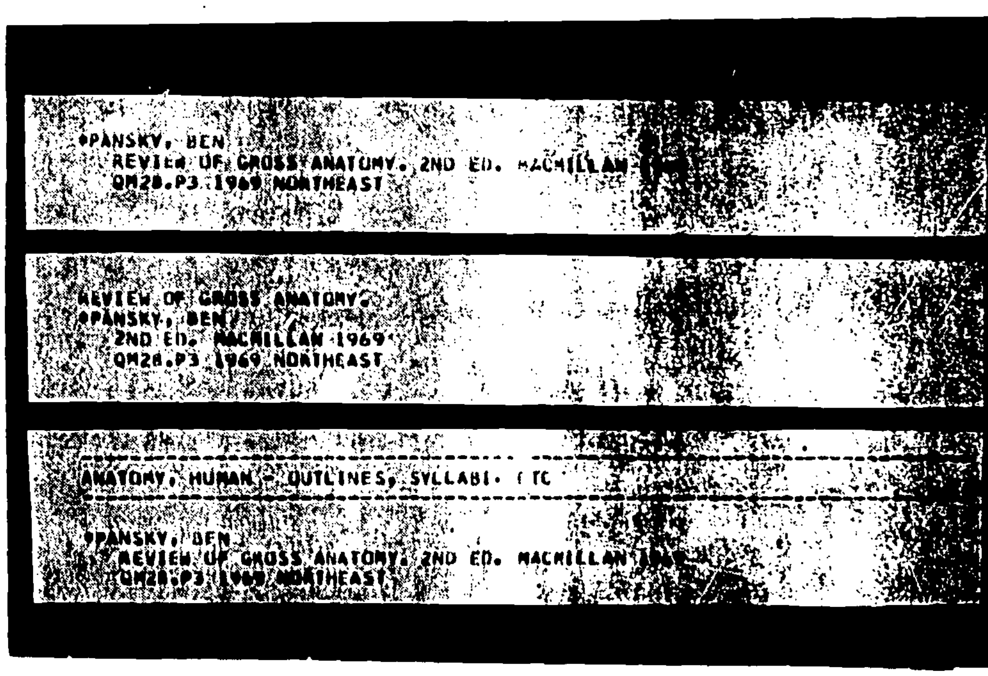
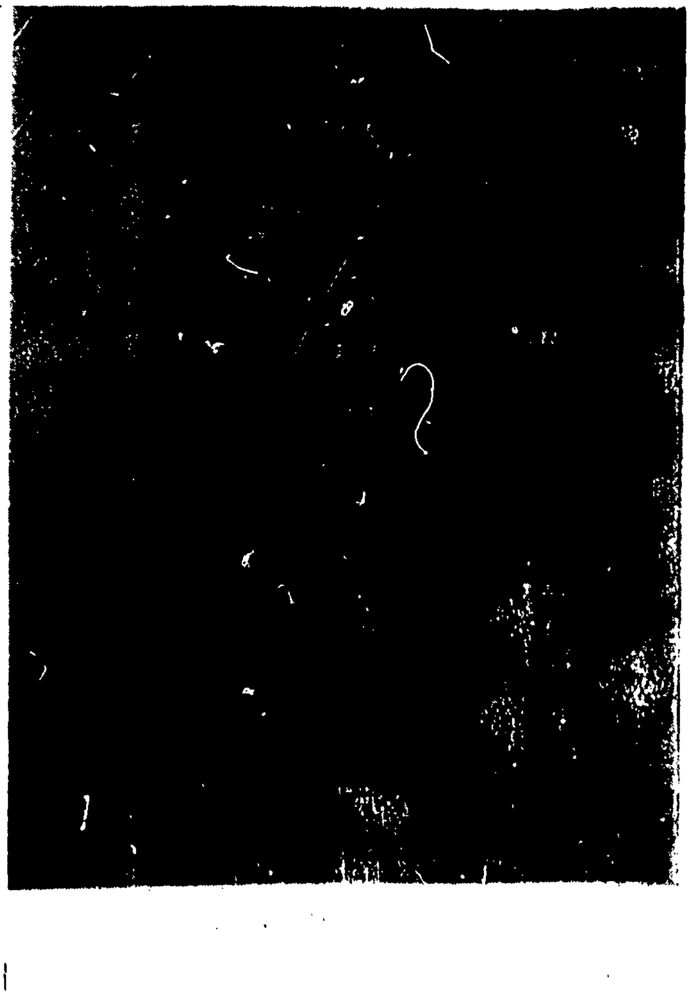


Figure 2

In finally deciding whether or not to utilize computer technology for the production of library catalogs, the most significant factors to consider are cost and appearance. This portion of the paper will be concerned with estimating some actual printing and production costs of three methods of catalog production. The costs pertaining to data conversion, storage and processing are not included in the figures nor are any costs pertaining to systems development of programming. All costs will be estimated from that point when actual printing begins.

Computer time on a machine large enough to handle the records of a library with 100,000 volumes or more rents for approximately $100 per hour. The line printers most generally attached to these computers are rated as speeds of 1,100 lines per minute. A more realistic output speed is perhaps 900 lines per minute. The average abbreviated bibliographic entry including the blank line between entries is composed of five lines of print. On that basis the cost of the catalog is approximately .00945 cents per entry. A library with 100,000 volumes incurs an expense of $945 to print one copy of one catalog. This figure is increased by a factor of three to $2,835 which represents the computer printing costs for one copy each of the author, title and subject catalog.

In a realistic library situation the public cannot be expected to have access to only one book catalog. The computer printing of 100,000 entries in each of three catalogs with an average of twelve entries per page produces approximately 25,000 pages of print-out. No one wants to thumb through 25,000 pages of print-out. Figure 3 is a sample of one page of a computer printed catalog.



In order to overcome both of these problems, the next step in library catalog production in book form is to hire a commercial printer to photo-reduce the computer print-out by as much as fifty percent and to produce multiple copies. These costs vary, but an average working cost is $2.00 per page for the finished catalog, bound in 150 copies. Using these figures and a total of 6,250 finished pages, the photo-reduction and duplication of the catalog costs approximately $12,500. This cost plus the $2,835 for computer printing produces a total of $15,335.

This represents a significant amount of money. The point that should be kept in mind, though, is that the catalog has been reproduced 150 times and is now capable of being accessed by not less than 150 simultaneous users.

Another method of reproducing library catalogs is through the utilization of microtechnology. The library catalog can be converted quickly and easily to microfilm and then duplicated for greater user accessability. Rotary camera microfilm machines are readily available at service bureaus or from manufacturers for nominal costs.

The most common reduction ratio for these microfilm machines is 24:1. The original 25,000 computer produced pages of catalog represents a mass of 22,917 feet of paper. Using the microphotography reduction ratio of 24:1, this mass can be reduced to 955 feet of microfilm which is a considerable saving in space.

Unexposed microfilm for use in the microfilm equipment cost $30 for each 200 foot roll. This price includes processing and developing by a silver halide process to produce a negative exposure master copy. The total cost for the 955 feet required for the library catalog then is $150 for one negative master copy. Duplicate copies can be produced either negative or positive, for an average cost of .046 cents per foot or $44 per complete catalog copy. To produce the same 150 copies as before, the cost now is $6,600. This is approximately one half the cost of the book form library catalog. Figure 4 is an illustration of a library catalog produced by these methods.

Figure 4

The third method of library catalog production utilizes a relatively recent tech-nological development called COM (Computer Output Microfilm). This technology has been developed within the last five years and should be of considerable interest to librarians concerned with producing library catalogs. A COM unit is an attachment tc a computer or a unit that acts as an auxiliary piece of equipment to a computer. Th COM unit acts just like a line printer with the exception that it prints directly onto microfilm at speeds far beyond any existing line printer. Currently operating COM units print at speeds up to 10,000 lines per minute.

The computer printing costs of $2,835 computed for the line printer can be reduc to approximately $300 by using a COM printer. The potential of COM for printing lar groups of data is obvious. It is something that libraries concerned with printing book catalogs must take into consideration.

All costs remaining equal, the total production figures for the three methods of library catalog production are: (1) book form, $15,335 (2) rotary camera microfilm, $9,435 and (3) COM produced microfilm, $6,900.

In conclusion, the production of library catalogs has been shown to be a valid concept from both a practical and an economic standpoint. Processing techniques, storage mediums and output technology has reached a level of sophistication that de-mands attention from library administrators. Through the use of modern computer techniques, librarians may come closer to realizing a library catalog that is accura complete, easy to update and economical to maintain.

Session Nine - Discussions

PROCESSING FOR AUTOMATION - I

Chairman: Prof. A. Beged-Dov (Israel)

MR. E. SCHAFER (US): In the WEIS project, essentially what we are doing is studying the interactions of the nations of the world. We have within our files 160 nations, each of them reacting or interacting with any one of the other 159. We have altogether 25,000 possible interactions. Some of these interactions are from zero level. The project is operational although it is a prototype. By prototype, I mean that it is of limited boundary. We select material from essentially one newspaper, *The New York Times*, but have also taken data from such papers as *The London Times*, *The Jerusalem Post*, *The Middle East Journal*, *The Bangkok Times* and so on. We cover about 25 to 35 newspapers, but due to limited funds we essentially confine our coverage to the articles in *The New York Times*.

The articles are scanned by graduate students of the School of International Relations at the University of Southern California. Two graduate students scan and code each article, so after a period of training, the coding is fairly consistent. The data is presently entered on-line in a terminal which is located in the School of International Relations. We built a data set which is sufficient in size to contain the entries for about a month (800-900 entries) of *The New York Times*. The coding is validated at the terminal as it is entered. In addition, there is another program for batch entry of deletions and corrections and these are entered in-batch in another program which also validates all these codings.

In the file structure, there is a zero level or basic level of the segment which is of a fixed length containing all the analytic code fields. There are subordinate groups of segments. You can visualize these as arrays. The first array contains up to 10 lines of abstracts. The other array, which because of financial limitations isn't employed right now, is to contain a full text of the article. In that we can include several hundred lines, up to almost 1,000 lines of 62 characters each.

We have used the GIS and other statistical programs written by graduate students in the School of International Relations. The students come to the School without any background in data processing. After a few hours of training, we have them writing programs for information retrieval using the generalized information system. However, this involves a compilation which costs money and time.

Last month, I completed a PLI program through which people can put in the parameters of the search. As a result the cost of the search is between an eight and a tenth of the cost of the GIS, because we are dealing with an object of marginal cost.

In addition, we have incorporated the possibility of searching on the text so that if you want to find out what Abba Eban said during a certain time period, or in the context of certain discussions between groups of countries, this is possible. Any number of searches may be entered at a given time. We are operating an information retrieval program, which enables us to find out, not what has been done, but what is happening now, so as to be able to make intelligent decisions based on that information. We want to do it in the context of interactions between nations of the world, to find out what the patterns are and where the patterns are changing and who is doing what to whom.

It is not enough to be able to get a list of abstracts, in a chronological or inverse chronological order.. What we want to find out is the significance of what is going on, and this is rather difficult because political and social sciences are notoriously soft. How do we know that the data incorporated is correct? We don't. All we can put in is what we read in the papers.

Since the data is soft, we have developed a series of programs called WEIS (World Events/Interaction Survey) Process to interpret it. The WEIS Process examines the data by various means, first by volume. We want to find out the national in-put output parameters. Second, the differential behavior changes. This computes the relative uncertainty index. Third, it examines the structure of relationships between countries and how they are changing, like the weather map. The WEIS Process monitor is used continuously every time we up-date the file, which is done about every two weeks. We construct the Main file and also an Analytic file, and every time the current up-date is entered into the file, the data in it is subject to this examination. We can see if the updating has anything unusual and where the patterns are changing.

We have fixed some experimental thresholds, the output of countries, and we want to see if the in-put exceeds any of the experimental thresholds. The variety indicator is sensitive to bi-lateral relationships. We build arrays of these dynamic relationships and throw out the ones that have zero in them, and then we eliminate data that have low usage and will result in no output. The result is about 100 bi-lateral candidates which are passed through subsequent screens where the output can be plotted. The first is called INTSCAN I (International Scan). We can set the control parameters up or down, and these give us a summary on specific events.

Next we trace the linkages between the countries. For this purpose we are attempting to use graph theory. If the volume and variety correlate, this can be used to buil a triad. We have a theory which gives some expectation parameters. We want to know wh the third parties are and how they are involved in any relationship. We have assigned sign values from the events coded.

We have nine different ways to go into this matter. Semantic differentials are used to generate a valance matrix and to reveal triadic relationships. After monitoring the data, we can use a time series analysis to examine what is going on in our files and perhaps to predict to some extent what the interactions are going to be in the relatively near future. We need to construct a model for computing simulations in order to build a theory on what is going to happen. We are presently building two sub-system models to determine the amount and type of behavior of the countries; for example, the reciprocity relationships between the US and USSR. We try to determine some fundamental theories. We have two right now that we are looking at: the reciprocity theory of behavior and crisis theory and the shape and form; and the time series analysis using what we call domain analysis in which we are looking at the time, the frequency and the amptitude. It is a kind of spectral analysis. Each event is like a channel and results in a sort of inter-channel analysis. Actually we are learning a good deal from electrical engineering.

As I said, the program is operational. We have some interesting observations and hopefully in another year or two, we will be on firmer ground.

MR. S. ISAACSON (Israel): In 1958 to 1962, a pioneering team working under the auspice of a military organization in the US, developed some new techniques of information tran fer. One of the techniques we developed in 1958 was the SDI concept in conjunction wit

IBM. The other was an attempt at automatic abstracting. We took terms which we used in indexing and strung them together into telegraphic abstracts. In 1960 we were able to produce abstracts at about 39 cents each.

MRS. T. KRIEGER (US): Most of the papers delivered this week have dealt with the information contained in documents. Most of the documents, however, are in the form of serials. The tremendous amount of serials makes their control very difficult and locating them is often a matter of chance. Most libraries do their cataloging independent of one another, and relevant to the nature of their own particular collection, though in the US the Library of Congress does most of the cataloging. Much discussion has taken place both in the US and in Europe on how to control serials; how much control do we want to have; what data elements are important; can one do this on a computer? If so, who is going to do it and who is going to make the decisions on what data elements should be included and in what order? These discussions began many years ago but the first formal study was made by the Information Dynamics Corporation in 1965.

Our study was started in 1969 by Mr. A. Lebowitz, who is now at the Technion in Haifa, and Mr. Joe Sheraton, who is the director of the National Agriculture Library. The three national libraries in the States, the National Agriculture Library, the National Library of Medicine, and the Library of Congress began to cooperate to solve some of these problems. As these libraries do not have centralized processing of any kind becsuse each is under a different department, the National Serials Pilot Project was started.

The two big decisions which were made fairly early were that the MARC format for serials, developed by LC, would be the accepted format, and that the Union List of Scientific Serials in Canadian Libraries would be used. The start was made for sci-tech materials, since money is more readily available for sci-tech than for humanities or social sciences.

The objectives of the project are listed in my paper. Objective two - producing a list of serials not held by any of the national libraries, was not attainable, because this had to be based on a list of serials of the world and to date no one knows what their number is. We have seen figures anywhere from 100,000 to a half a million live serials.

There is a great deal of disagreement as to what a serial is. Conference papers are treated as serials by one library and as monographs by another. The data program decided that a serial will be what anybody treats as a serial. Our machine problems were enormous since we were using two different computers. Our programs had been written in basic assembler language by the Library of Congress but we were not able to use their computer for processing. But the machine problems, great as they were, were easier to solve than the data problems. When we started there were 5 error messages. Now there are 270 machine detectable errors.

One of the basic problems was on entry. As I mentioned in the paper, there are two facets to the term. We were concerned with the ordering of entries. Is there a need in a machine file for a main entry, since one can pull out the material in any order they like?

One of the recommendations made in the final report was that this be re-examined, that one should discuss serials in the light of the nature of serials in an electronic data processing environment.

Various universities in the States received grants from different sources to work out some of the serial problems. They have found much the same things as we. When we started, we read many of these reports and found they didn't apply to us. On re-reading them, we found that they really did apply to us. We were looking at the differences rather than the similarities. At Colorado Richard Doharty, who attempted to borrow a tape from another library and to alter it for the conditions at Colorado, makes it very clear in his paper.

Some of our problems seem minor, e.g. modified letters. We decided to eliminate them completely. If anyone has any ideas on that topic, I am certain that the people at the LC now working on the project would be delighted to know about it. The ALA has recently come up with a standard of 126 characters. Whether we will use them all or a subset of them, I don't know.

The MARC system allows everyone to dream about putting everything into the system. But if you put in everything, you don't have time to put in many different things. We felt that if some information was contained in a fixed field, such as the beginning date, there was no reason to repeat it in the body. The MARC format was designed by the LC to produce catalogue cards, therefore, there is a lot of redundancy built into it for their particular purpose.

We did not use Standard Serial Numbers. We used a Local System Number to see if we could actually link records together by these means. We came to the conclusion that the up-dating procedures would have to be done at the same time as the base was developed because everything kept changing. Change seems to be the most common characteristic of serials.

We did not use much statistics as our file is rather unique. It is random only in the sense that it was not planned. It is not a random selection of serials. It contains 900 items from the National Agriculture Library, 2,500 items from the National Library of Medicine and the rest is A to Z of the Canadian list. The idea was to get it into a base to see what would happen when we try to manipulate it. The file has 7,059 items, of which 6,221 are live. 3,800 started publication since 1946. In 800 cases the publisher was the main entry.

When we started we put in everything with as much redundancy as the system allowed. Later we eliminated much of the redundancy and now we have a mean average character record of 420 and mean average number of fields of 9. Working in three libraries, one basic problem was whose entry to take. Originally we took that of the LC but this didn't work. We decided to take that one which somebody gives us first. But in order to reconcile the variations we are strongly recommending that an independent authority file be set up. Without that, you must know precisely what the item was called in order to retrieve it. Publications from 104 countries are included. Initially we used a graphic code, based on political boundaries. This did not prove itself. Now a code on geographic divisions has been developed which I hope will be more stable than political boundaries.

We are aware that many of our recommendations are not feasible economically, nor will they be greeted with joy by some of the libraries who must think not only of serials but how serials fit into their holdings.

CHAIRMAN (Prof. Beged-Dov): I will take this opportunity to mention another aspect. It is not always necessary to understand the information so long as you get the right reaction. A few years ago when we went to the States, my son was 7 years old. He went to first grade and he had quite a few disciplinary problems. One day he came back home saying that he could not go back to school without either his mother or father. My wife went and the teacher explained the problem. My wife did not understand her English so she turned to my son and asked, "Rafi, what does she say?" To which he replied, "Mother, what does it matter; start crying!"

Information science is in a state which certainly requires a very capable cadre of technicians. It is certainly not a glamourous work and the glamour associated with it wears out very soon and it takes perseverance to do it.

QUESTION: *To Mr. Schafer.* Where do you get *The New York Times* tapes and the use of other data bases? The larger the data base the better your intelligence is going to be. Though you use primarily *The New York Times*, it is a tremendous data base.

MR. SCHAFER: We have been negotiating with *The Times* to use their tapes on a reciprocal basis. Because of the economic considerations, we aren't taking data from other newspapers now. We have done some character studies of reports reported by two or three other papers to see what differences would result in the statistical program. There are some differences, but we haven't found them really significant. So we are confining our work to *The Times* right now. This is a prototype arrangement. Hopefully, in time, we will have established that the project is feasible and will then have more funds, and be able to expand the coverage.

QUESTION: *To Mrs. Krieger.* I wonder if you agree with Dr. Garfield's position that a few hundred journals contain most of the information. If you do, how do you justify the investments in serial control and if not, what is the benefit of control in relation to cost?

MRS. KRIEGER: I don't know how true Dr. Garfield's statement is. I do know that there is a lot of duplication and we see it fairly often when we examine the journals. But I don't know which journals are the important ones. This is a matter for each librarian to decide. There have been studies made on quality of journals, but this is not part of our interest.

You asked about the cost of controlling serials. At present almost every large university is starting a serials project. The projects are of two types. One is a union list, a list of the library's holdings for the benefit of other libraries, and the other is an internal housekeeping program to take care of acquisitions, binding, and the location of serials within their own university. All the systems which are being developed are suited only to a particular situation. This seems like a tremendous waste of time and money. If there were one central place where you could get a tape which identifies every serial, and could use it to make your union list, it would be easier to interface with other libraries.

QUESTION: *To Mrs. Krieger.* As you perhaps know, UNESCO is trying to establish an International Serials Data System. This system is based on the project of Mr. M.D. Martin of INSPEC and will be located in France. It will be composed of local and regional centers and the "international center". In the Working Group we have already discussed basic data elements. Our project will be compatible with the "Standard Serial Number" system. I hope that we can cooperate on the national level with your project.

MRS. KRIEGER: We are aware of the Martin project and we hope that there will be a lot of cooperation; primarily that we won't all be entering the same titles at the same time.

MR. D.G. KINGWILL (South Africa): In South Africa we have been using a native system; we mechanized our union list of periodicals using an IBM print check, and we use diacriticals, because the major language in South Africa is Afrikaans which uses them.

With regard to the question about the number of journals which are useful, I think in London the number is something like 150 and the number of titles is something like 36,000. The catalogue is the main difficulty. Many publications are published under official names and many countries have changed their names, so the life history of journals is quite a difficult matter. The Canadian system with some adaption has met our requirements very well.

With the computerized system we are able continually to update, although the published version takes some time. On the Telex, we are dealing with 500 to 600 inquiries a month for location of journals.

QUESTION: To Dr. Kessler. I got the impression that after the first initial trial period, the bulk of your inquiries could be handled through a batching process. Is this the case?

DR. M. KESSLER (US): The answer is yes and no. It could be used in a batch mode. But then it would become a terminal type of thing and would have no possibility for evolution. By leaving it in the interactive mode, with a sort of pseudo-batch possibility, in the sense that the user could instruct the computer to merely store the inquiry and put it out once a month when new data comes in, it gives the user the opportunity to experiment and interact with his profile statement; and change and modify it, and measure its effectiveness against the response. That can be done on-line. The other possibility lies in the various forms that I described, like leaving it in his private memory, or having it sorted in various ways. For example, some people like their information sorted by journals; others like it chronologically or alphabetically by author. There are a considerable number of options left to the user in having it on-line, although it is true that one of the options is to treat it as if it is a batch.

MRS. A.M. DE BUSTAMANTE (Mexico): In the light of Mr. Kessler's experience it is advisable to review some of our points of view. If I understood correctly, MIT created a file for physics literature searching that nobody is using the way it was planned. However, some of us are planning to create files, among other things, to do retroactive search. Dr. Garfield believes that we have to keep our files for searching. Yet one of the more advanced research centers at MIT does no retroactive searching, at least not for physics. They use the SDI and create their own files.

QUESTION: To Mrs. Krieger. In considering building the serial catalogue, was there any plan for changing the card system?

MRS. KRIEGER: There is no intention to eliminate the cards. They can be put into books. What we are really talking about is putting the data into a computer and being able to publish books in various ways, such as new serial titles. We are not talking about eliminating catalogues, only about offering the information in such a manner that it can be used in any way suitable for users' particular needs.

MR. U. BLOCH (Israel): To Dr. Kessler. I understood that Mrs. Krieger wanted to have a tape and the idea is that you do not need cards or books. You have everything in the computer, and can search for any field without having to have different cards catalogues for the different entries you want to look for. The above also has some relevance to your problems because the Pilot Project will be another data base that would worry you for your integrated system. I don't see why you are worried if an additional abstract service, e.g. on ecology, is added. It means that instead of running through 17 services you have to run through only one, the one on the subject of ecology. I don't see why it worries you unless different services are differently structured.

Also, to what extent did you try to indoctrinate your physicists to use the system interactively?

<u>DR. KESSLER</u>: *To Mr. Bloch.* Why do I worry about having yet another tape on ecology? As tapes multiply the housekeeping problem within an information system becomes greater; it becomes a library in itself. There is the handling of tapes, their storing etc. Aside from that, what worries me is that abstract journals depend on people's opinion as to how they should be indexed. Who are the people in Raleigh, North Carolina to tell me what in CA is related to ecology? The abstracts are already one or two phases removed from the author and now we are introducing another middle man. How long can this go on before it becomes completely divorced from the author's intent?

Also, there is the factor of endless multiplicity; e.g. we have a department of urban problems. What is the data base for urban problems? As we get more and more into multi-disciplinary problems, shall we just keep creating tapes?

As to your question of whether we tried to convince the user that it is worth his while to use our services, here again, how do I know that it is worthwhile? We are talking about the on-line interaction. It may be worthwhile in some cases and not worthwhile in other cases. One ought to leave the options to the user and through the evolving process of education, let him decide when it is worthwhile and when it isn't. We held several seminars with the physics faculty and graduate students. We have distributed printed educational material. We installed two consoles in the physics department with experts present who were available for consultation. I don't want to overstress this. We merely gave them options. Some people do interact for a few months and become trained in the use of the process and then find they have no need for it and they don't and we don't urge them.

SESSION TEN

processing for automation–2

Authors of Papers

THE SYSTEME P.A.S.C.A.L.

J.H. d'OLIER
Directeur-adjoint du
Centre de Documentation du CNRS
15 quai A. France, PARIS 7e .

In order to attain maximum efficiency (speed, precision, adaptability to users' needs) in scientific and technical information transfer, the Centre de Documentation of the CNRS (French National Centre of Scientific Research) has developed and is now operating system P.A.S.C.A.L. (Programme Appliqué à la Sélection et à la Compilation Automatiques de la Littérature).

Source Materials

9.000 journals, Conference proceedings, theses, reports, etc., are regularly scanned and analysed.

Each abstract (500.000 per year) is given a classification code (or more if necessary) and then indexed; bibliographic references are strictly controlled. The basic work-sheet, on which have been entered the complete bibliographic record, translation of the original title, short indicative abstract, key-words and relevant codes, is sent to be punched for processing in the machine system, the final object being multiple entrance to all information stored.

Methods

Input - Originel documents are processed as soon as received. The work-sheets are sent of keyboard operators for introducing into computer and printing by typographic methods which give more facilities for corrections than punching cards. From the codes and key-words so entered, the computer sorts the abstracts according to a classification established in the beginning of the year; it generates also author indexes in conformity with ISO rules and subject indexes with the key-words written on the work-sheets when indexed.

Finally, the magnetic tapes are transformed by suppressing typographic codes and SDI services are performed. Three groups of files are processed as follows :

- One is an inverse file in which the key-words are classified in alphabetic order and followed only by the document numeros
- The second is a direct file which contains the complete bibliographic descriptions and abstracts arranged by their numeros
- The third file contains the SDI profiles.

Services afforded to users :

Besides the thirty sections of Bulletin Signalétique, monthly SDI services on personal or standard profiles (designed in collaboration with users) are possible in all fields covered by mechanized sections (Polymers - Pharmacology - Microbiology and Virology - Information Science -Astronomy and Geophysics - Earth Sciences - Pathology - Genetics - Psychology - Electricity - Nuclear Physics, Chemistry and Technology - Energy - Metallurgy - Applied Chemistry - Water and Air Pollution).

Retrospective searching

This service will be undertaken with computer in 1972 when one year of documents will be accumulated on the magnetic tapes. As basic time for such services is five years, the CNRS will give answers partly obtained from computers for the more recent documents and partly retrieved by classical method.

After one year, the contents of the direct files with be transfered on a photographic memory because the processing on computer files is considered as not necessary and too much expensive. However, we are also developing a retrieval system based on the natural language used in the titles and in the bastracts, which will complete the retrieval system by key-words. In the sections in which that study will succeed, a part of that direct file will be kept in the computer and included in the inverse file.

SOME STUDIES ON COMPUTERISED INFORMATION RETRIEVAL TECHNIQUES BASED ON CA CONDENSATES

F.H. Barker, W.E. Batten, D.C. Veal, B.K. Wyatt.

United Kingdom Chemical Information Service,
University of Nottingham, University Park,
Nottingham NG7 2RD, England.

SYNOPSIS

An evaluation has been carried out based on the Chemical Abstracts Service magnetic tape data base CA Condensates. The study has investigated the relative merits of searching titles, titles plus keywords and titles plus abstracts. Factors affecting profile performance and methods of automatically constructing profiles for natural language searching have also been studied.

Introduction

At the beginning of 1968 Chemical Abstracts Service (CAS) announced a new magnetic-tape service, <u>CA Condensates</u> (CAC). This service covers all the items selected for <u>Chemical Abstracts</u> itself, and contains for each item the title and usual bibliographic data, enriched by keywords. Since this service made available the total CAS data base for the first time in computer-readable form it was an obvious target for detailed study. UKCIS began an experimental project on this data-base early in 1969, under the title of the CAC Evaluation.

Objectives

The CAC Evaluation had the following objectives:

(a) performance correlation

Earlier UKCIS work (3) had investigated the extent to which profile performance is affected by such factors as subject interest, type of user, search logic, etc. The CAC Evaluation continued this work and attempted to overcome some of the methodological problems found earlier.

(b) data-base comparison

The advent of CAC meant that three types of data base were provided by CAS: - titles only, e.g. Chemical Titles (CT)
- titles plus keywords. e.g. CA Condensates (CAC)
- titles plus abstracts, e.g. Chemical-Biological Activities (CBAC)

These three types differ in having a different number of access points per item, i.e. a different number of words available for searching, by which retrieval may be effected. As the number of access points per item rises, one can expect recall to increase, precision to decrease, and the cost of searching to increase. Obviously, information on the relative cost-effectiveness of these three approaches is of great interest. One of the major aims of the CAC Evaluation was therefore to compare the three in terms of cost, retrieval performance, and also currency.

(c) automation of profile construction

The CAS data bases are generally constructed in natural language, that is to say, no control is applied to the vocabulary used (although foreign language items are translated and there are some rules governing the preparation of keyword phrases). This means that search profiles to operate on these data bases must take into account grammatical and syntactic variations, synonymity, broad-term/narrow-term relationships, and so on. The construction of adequate profiles can thus present severe problems, particularly to naive users. In the past one of the most useful ways to overcome these problems has proved to be careful study of papers already known to be of interest to the topic in question. Such a study can be most valuable in indicating the variations in terminology and phraseology which should be covered by the profile. The value of this exercise increases with the number of papers studied; so, unfortunately, does the tedium. Hence, in the CAC Evaluation UKCIS has been studying the feasibility of automating the analysis of known relevant items. The long-term aim of this work is to reach a stage at which it is only necessary for the user to indicate whether items presented to him are of interest or not. By analysis of these decisions the computer will 'learn' what sort of papers the user is interested in and will itself devise the best way of selecting them. In other words the computer will write the search profile. A more detailed description of the procedure has been given elsewhere (2).

(d) general

In addition to the above major objectives, several other aspects were investigated. These included the value of new logical facilities introduced by UKCIS at the start of the experiment and the distribution of relevant items over the 80 sections of CA.

General Outline

The first stage of the project was the establishment of a user population. This was done by selecting some 200 profiles from 45 different organisations in the U.K. The profiles were carefully selected to cover a range of subject areas and types of interest. The U.K. Universities were not generally included in the selection because they were already extensively involved in other evaluation studies (4). All the profiles were searched against CAC, and appropriate subsets were also searched against CBAC and POST (Polymer Science & Technology) for the data-base comparison. The searches were carried out in three phases.

In the first phase users received a standard current-awareness service. This was to enable them to develop profiles in a normal operational environment which could then be used as a basis for comparison with modified profiles developed later.

In the second phase repeated 'iterative' searches were made against fixed files of information consisting of three issues of each data base (including both three odd-numbered issues of CAC and three even-numbered issues.)

In the third phase the results of the fixed-file work were checked by providing current-awareness searches using competitive profiles for each topic.

Performance Correlation

The user, questions, and profiles are being analysed in accordance with a detailed classification scheme developed earlier (1). Some preliminary results about the user population are given in Table 1, and Table 2 shows the subject areas covered by the profiles, in terms of the five CA Section Groupings. Note that Table 2 contains some overlap since 57 of the profiles involved more than one Section Grouping. The correlation of these and other factors with profile performance is now being carried out, using standard statistical methods.

User Population

	Industrial	Government	Other	Total
Organisations	33	6	6	45
Users	110	20	12	142
Profiles	146	25	22	193

Table 1

Subject Areas Covered by Profiles

	Biochem	Organic	Macromol	Applied Chem & Chem. Eng.	Physical & Analytical	Fringe
No. of profiles	54	58	44	15	71	9
Percentage	28	30	23	8	37	5

Table 2

Data-base Comparison

This study involved 19 profiles which were suitable for searching against POST, and 29 suitable for searching against CBAC, in addition to CAC. The users were given the opportunity of developing, for each topic, profiles specifically adapted for searching titles alone, keywords alone, titles plus keywords, and titles plus abstracts. These profiles were then searched competitively against the fixed files, and the results are being compared. On CAC, the first three profiles were also used in the final current-awareness phase. Results from the latter are not yet complete, but some preliminary figures are given in Table 3.

Automation of Profile Construction

The profile developed during the initial current-awareness phase was searched against the fixed file, and the output was assessed by the user, to provide the initial batch of known relevant items for analysis. These items were analysed by computer and potentially useful search terms were selected on the basis of term frequency studies (2). The selected terms were automatically formulated into a profile which was again searched against the fixed file. Any new output was sent to the user for assessment and any new relevant items were added to the store for analysis. This procedure was repeated iteratively until no new relevant items were retrieved. On average this required about three iterations and retrieved an additional 20% of relevant material (thus showing that the recall of the original profile could not have been higher than about 80%).

Comparative Performance of Title and Keyword Searches

	Total Hits	Relevant Hits	Total Unique Relevant	Precision (%)	Recall (%)
Titles only	365	243	361	67	67
Keywords only	574	269	361	47	75
Titles plus keywords	844	332	361	39	92

(N.B. Results for 10 profiles)

Table 3

The study of the relative currency of the data bases has been completed, and the results are given in Table 4. All figures are percentages.

Delay (in weeks) Relative to Chemical Titles

	2	4	6	8	10	12	14	16	18	20
a) Journal Issues										
CBAC	0	2	22	31	19	14	7	4	-	-
CA (odd)	-	2	12	15	16	13	12	6	3	2
POST	-	0	7	17	20	17	16	12	6	6
CA (even)	0	13	20	21	18	13	6	4	2	1
b) Number of papers										
CBAC	0	2	30	33	17	10	5	2	-	-
CA (odd)	-	1	12	29	31	15	7	4	1	0
POST	-	0	5	15	19	18	22	11	6	3
CA (even)	2	28	29	18	12	6	2	2	2	1

Table 4

The final search term list was used to modify the the original profile and the original and modified profiles were run in competition during the final phase. Again results are not yet complete but some provisional figures are presented in Table 5.

Effect of Profile Modifications

	Total Hits	Relevant Hits	Total Unique Relevant	Precision (%)	Recall (%)
Original	3515	1314	1484	37	89
Modified	3102	1320	1484	43	89

(<u>N.B</u>. Results for 94 profiles)

Table 5

General

The new logical facilities developed by UKCIS included syntactical linking and IGNORE logic, a more flexible variation of NOT logic. After instruction and recommendations by UKCIS staff 17% of CBAC profiles, 36% of POST, and 37% of CAC, used the new facilities. The effect on profile performance is being studied.

The distribution of relevant and irrelevant items over the 80 CA sections has been investigated for each profile. These results have proved very useful since Section Numbers can be used as search terms.

Conclusions

Obviously much work remains to be done on this project and so it would be premature to try to draw definite conclusions at this stage. Nevertheless the database comparison is providing extremely valuable information for guiding the future development of computerised information services. Also, the preliminary results from automatic profile construction have been sufficiently encouraging in the limited studies so far carried out to justify extension of the analysis to cover term pairs, term triplets, etc.

References

1. Barker F.H., Analysis of Search Questions and Profiles, UKCIS, May 1969.
2. Barker F.H., Veal D.C., A Procedure for the Semi-Automatic Construction of Search Profiles, for a Mechanised Information Retrieval System, Chemical Society Autumn Meeting, Southampton, September, 1969.
3. Barker F.H., Kent A.K., Veal D.C., Report on the Evaluation of an Experimental Computer-Based Current Awareness Service for Chemists, UKCIS Research Report No.1 London, 1970.
4. Leggate P., University of Oxford Experimental Information Unit, unpublished work.

CHEMICAL CODES IN INFORMATION RETRIEVAL

E. Hoffmann
National Center of Scientific and Technological Information, Tel Aviv
and
Department of Organic Chemistry The Hebrew University, Jerusalem

SYNOPSIS

Chemical codes are an instrument for gathering and refining chemical information. The manipulation of chemical data and information by computers necessitates the change of chemical codes into alphanumeric chemical codes which can be divided into 3 broad categories. Most of the development of these codes has been carried out by industrial concerns for well defined purposes and therefore a multiplicity of such codes - each with different specific characteristics - has sprung up. Thus chemical information stores with different alphanumeric chemical codes are not easily compatible. However, on the other hand, codes exist and can be adjusted to satisfy every need of any specific chemical information system.

The subject of my talk covers a very wide area of interest and activity and it seems unwise, in such a short talk, to try and give a comprehensive overall view of the development in this field. Therefore I shall limit myself to the elaboration of specific points and shall assume that the audience is more or less familiar with the pertinent background information.

Chemical codes in information retrieval sound suspiciously like mechanized chemical information systems and the very first point I want to make - and emphatically so - is, that this need not necessarily be so. The truth of the matter is, that chemical codes as well as chemical information retrieval were practised in chemistry long before the advent of any mechanized information system; simply because any systematized type of chemical communication or expression falls quite naturally under this category of codes. Thus any type of nomenclature - systematic as well as free nomenclature, brutto formula, and finally structural formula is a chemical code with varying degrees of chemical information content. None of these sprang into existance, de novo and the development of the chemical sign and symbol language - or code, if you wish, as examplified today in the structural formula, runs a close parallel to the development of chemical knowledge, concepts and understanding. Indeed the structural formula is probably as basic as indispensible an ingredient in the development of chemistry as mathematical methods are for the development of physics. Here in chemistry, as well as in physics, a special medium is necessary in which the subject knowledge can be embedded, processed, reshaped and worked over. Indeed the whole conceptual development process is carried out within this matrix - which is the chemical structural formula in chemistry and mathematics for physics. One of the most pertinent and widely quoted examples of this interaction between the code language and the subject matter - in chemistry - is the development of the benzene formula by Kekule. Here, the substitution of a specific structural pictural chemical code was able to solve the dilemna of the unsaturation of the C_6H_6 compound. The true interaction between the development of - the chemical sign language on the one hand - and the basic chemical understanding and knowledge on the other hand - comes here to the fore. Thus the chemical structural formula represents a chemical code - par excellence - and an extremely useful and valuable one. Indeed our everyday chemical thinking operates in structural formula terms and a large part of the chemical user community namely the organic chemical industry is inter alia concerned with chemical structural problems as represented by the chemical structural.formula, and is vitally interested in retrieving chemical structural information.

Thus far, therefore, chemistry has done very well indeed with the structural formula as a code for dissemination of chemical information, but two new developments tend to impinge on this situation and a consequent change may be in the offering.

First; new basic chemical knowledge is being obtained which can be represented only with difficulty with the chemical structural formula in its present two dimensional shape. Thus in the meantime a host of auxiliary descriptive devices have cropped up especially for stereochemical information such as the Newman projection formula, the boat and chair puckered form of a hexagon, the Cahn-Ingold-Prelog system for stereochemical classification, the explicit notification of three dimensional structure by bold and light face dashes. In addition there is a host of new electronic bonding information available which cannot presently be accomodated within the structural formula as drawn today. And so one is on the lookout for more appropriate pictural representation of the accumulated chemical electronic knowledge.

Second; the amount of classical structural information - as represented by structural formulas - available grows at such a rate that it cannot be usefully processed anymore without the help of computers. This means that a machine readable equivalent for the chemical formula has to be created. All machine readable and processable chemical codes use alphanumeric symbols and can roughly be classified into three broad categories.

A) Topological systems
B) Notation systems
C) Fragmentation codes

A) Topological systems in general represent a complete replica of the structural formula and its information content. Each atom within the molecular formula is represented by a node and its connection with adjacent notes are explicitly specified. In this way the whole structural picture is represented in a kind of numerical matrix (the nodes being numbered). The structural information is complete and all or any part of it can be retrieved or querried. The system is extremely simple to process for input into the computer and this is advantageous in so far as no specially qualified personell is needed for this operation. The computer store needed is however very large even with condensed canonical systems and this makes for the use of long computer times in search queries. Thus retrieval procedures from chemical data stores based on topological systems tend to be long and expensive especially where large stores are concerned. These systems tend to be non unique and unambiguous representations of the structural formula but they can also be unique if they are canonical.

B) Notation systems: These are a sort of short-hand scripture whereby symbols are substituted for atoms or groups of atoms within the molecule. They can constitute a unique and unambiguous representation of the structural formula, provided the notation rules are specific enough to make them so. They need to be learned in order to execute an input, and the personell need to be qualified to a certain extent in chemistry in order to choose the correct notation. They do not take too much storage space and are easily queried and do not use up much computer time but not all structure queries can be answered by notation systems.

C) Fragmentation Codes assign code designations or terms mainly to specific groups of atoms and bonds. The results are ambiguous codes that are not completely descriptive of the corresponding structure, i.e. a number of similar structures are covered by the same coding sequence. Computer storage is relatively compact and relatively easy to implement. Querries are answered fast but do give some irrelevant marginal material (which is sometimes considered an advantage if there is not too much of it).

Having thus characterized machine readable chemical codes one should at once point out that this broad classification into 3 so called different notation types is by no means absolute. And this is so for some very curious reasons. A closer scrutiny of the development of computer stored and manipulated chemical information systems will show that most of them were initiated and have been developed by industrial enterprises. If we notice that most of these are commercial, profit-making, enterprises (government agencies such as U.S. Army, Air Force, NASA etc. excluded) it must be supposed, interestingly enough, that the development of such systems for these enterprises was - and is - economically justifiable. This is a point which is still debated in information research - and research establishments. However each of these industrial enterprises established and developed their own information systems - and the notation system that goes with it - according to their own specific needs and interests, which are diverse and different. Thus a multiplicity of different kind of notation systems have been created not only in order to retrieve the absolute chemical information per se (which is not essentially their main task) but in a way and in a manner specified by the aims and purposes of the prospective user. Thus the differences between the various systems are sometimes not absolute ones, but conditioned ones and therefore tha arguments, which is the best notation system are really meaningless. Each system is and can and should be adopted - or created - according to the specific needs of the user.

Thus there do exist presently quita a number of independent chemical computerized information stores, each operated with its own developed system and using as input any of the existing notation methods or even a specific version of one. Thus the stores among themselves are very often incompatible in access and utilization of the stored information and even information stores which offer commercial tape services from their core are not always compatible - to the chagrin of the user. Thus very early in the game, the problem erose to make the various input methods - and with them, of course, the so recorded information - interconvertible and interchangeable by the superposition of special conversion programs. Much effort has been invested in these endeavours and their aim and purpose has generally been in two directions.

1) Conversion into topological systems: Since the topological system is generally an exact and all encompassing equivalent of the structural formula and therefore

carries all its chemical information content, the effort has been made to transform less comprehensive notation systems such as linear notation systems and fragmentation codes into topological system and thus be in a position to answer structural querries which the store would have been otherwise unable to answer. Thus I.C.I. has developed their CROSSBOW system which transforms their WLN stored formulas into connectivity matrixes with consequent enhancement of their retrieval and answering capabilities. Similar efforts have been made by DOW and N.B.S. and others, mainly effecting the transfer of WLN into connectivity tables.

2) Conversion from topological systems: These efforts are mainly concerned with the largest chemical computerized information store, i.e. the C.A.S. This store uses connectivity matrixes as storage medium and comprises presently over 3,000,000 compounds with the file growing at the rate of 5,000 compounds/week. The enormous size of this file makes it a very a valuable one but its topological structure makes a serial search in it prohibitively expensive. Thus quite a number of efforts have been initiated to transform this file into other storage media, line notations or fragmentation codes, which lend themselves to computerized search in a much more economical fashion. This type of transfer seems to be the more difficult one to program - according to Lynch[1] - but recently a programmed automatic method for transforming topologically stored formulas into WLN notations has been published[2] A similar capability for transformation into the Gremas system does exist[3].

Sometimes the various different properties of the different chemical codes have been put to complementary use within one chemical information system. Thus I.C.I. effect the input into their data store by encoding new chemicals according to the WLN. These are then stored in the computer in WLN notations (or their equivalent). I.C.I.'s experience is that encoding is very fast and efficient and the computer storage is compact and economical. Searches on this store are carried out in its original form whenever possible. A computer program exists which can transform the WLN into a topological code system (CROSSBOW). This is being used for checking purposes (the mol formula can be calculated from the topological matrix and compared = checked with the originally imputted mol formula) and for those querries that cannot be handled by WLN. The topological matrix is not stored but is erased after use and reproduced again whenever necessary. A different combination of codes is being practiced by the I.D.C. The I.D.C. uses the GREMAS system for its retrieval operations. However the chemical formula is encoded according to a topological system because this is the method which necessitates the least qualification for execution. This stored topological matrix is then transformed by a special computer program into the GREMAS code which is then utilized for search queries. Thus input is effected as easily and as economically as possible and the same applies for output.

Another chemical code - a fragmentation system of special type was devised by the "Dokumentationsring der pharmazeutischen Industrie" and is called the Ringdoc system. The elementary dissemination unit of this system is not necessarily a magnetic tape or disk storage but rather a punch card. Thus this unit can be used as feed-in into various computers as well as manually operated search and retrieval systems. Every basic conceptual unit in chemistry is then represented by one specific hole - or bit - of the 960 available on the punch card. Thus just as in any fragmentation system the compound to be encoded is dissected into a finite number of basic units, each of which is punched on the punch card. Thus the totality of punches on it represents the compound. The same procedure is carried out adjacently for the corresponding pharmaceutical and therapeutical properties of the compound. In this way a complete picture of the biological - chemical activity correlation is punched on one - or if necessary several connected punch cards to be read by a variety of computers as well as by manual devises.

References

1. M.F. Lynch, Jour. Chem. Doc., 8(3), 130-33 (1968).
2. D.A. Koniver et alia, Jour. Chem. Doc., 10(1) 52 (1971).
3. A. Kolb, private communication

MACHINE-AIDED INDEXING AND ANALYSIS FOR DOCUMENT AND FACT RETRIEVAL

Walter E. A. AXHAUSEN and Andrew E. WESSEL

SYNOPSIS

A program system both for semi-automatic indexing and for searching literature is outlined. It operates in the conversational mode via a data display terminal. A laboratory version runs since May, 1970. Current improvements include the switching from document to fact retrieval. Core memory space is less than 130 kb.

The utility of automated retrieval systems depends upon achieving realistic solutions to the problems of data analysis and preparation. One such solution, a system called PRIMAS[(x)], providing computerized techniques for direct use by human document indexers and analysts, has been developed. It is currently being tested, at the German Patent Office, Munich, Germany.

Its laboratory version - in operation since May, 1970 - is restricted to the use of one data terminal, and of small quantities of data. By the end of this year, however, a new version is scheduled to become operational which will be capable of operating with several data display units, and with large quantities of data.

PRIMAS relies on a set thesaurus which is maintained by an updating service. This limitation is not to be understood as a simplification of the program but rather is due to the fact that for many types of documentation, a fixed thesaurus is desirable (see EURATOM).

For the maintenance of the thesaurus - a task which is necessary if the documentation is extensive - every possible aid is given by the computer. PRIMAS logs internally all the requests made to the system and evaluates the material statistically. This will enable comparisons to be made concerning current use of the descriptors both as a function of indexing and as a function of searching, and conclusions to be drawn with respect to the thesaurus.

(x) = Programmpaket zum Rückgewinnen und Indexieren mit Maschinenhilfe (program system for machine-aided searching and ind.)

Development is focused mainly on the computer-aided input of documents. It is here that the major part of the work of documentation occurs together with the greatest likelihood of error so that the use of a computer for this task brings notable improvements.

For reasons of economy at the user's end, the space for bibliographic data etc. is restricted. Thus for 30'000 documents, one disc pack is sufficient, at 27'000'000 bytes per disc pack.

The PRIMAS indexing dialogue operates as follows:

Soon after the beginning, a check is made whether the document is an old or new one. Next, any needed sector of the thesaurus may be brought to the screen and therefrom, relevant descriptors will be assigned simply by putting a "+" before them using a keyboard. Wrong assignments would be erased again by a "-" sign.

As soon as a few descriptors have been assigned, or sometimes even as early as after having noted the first descriptor, the computer is able to propose other descriptors specially for the document at hand, doing so of its own accord. This function is guided by all documents previously indexed. Ours tests with the existing program have shown that in many cases, not less than 70 % of the proposed descriptors were relevant so that they were immediately assigned, by the analyst.

Moreover, the computer will be able to compare how far a document just indexed concords with other documents already in the system. If the degree of correspondence were high, the indexer would investigate any differences by reference to the original, thus completing her, or his, indexing job.

For actual indexing practice, the following dialogue is typical:

No.	SCREEN (automatic):	USER ANSWERS (using keyboard):
1.	"This is PRIMAS; do you want a printed record of this session ?";	"YES" or "NO";
2.	"The following modes of operation are possible: MACHINE-AIDED INDEXING.......AN, EDIT.........................DA, MACHINE-AIDED SEARCH.........RE; please select a mode of operation !"	"AN"
3.	"Number of document to be indexed ?"	"1251367"

No.	SCREEN (automatic)	USER ANSWERS (using keyboard):
4.	"The following descriptors characterize document number 1251367: No; this is a new document number; please return to computer by pressing button !"	Presses button;
5.	Shows thesaurus beginning with highest level of its hierarchy;	Chooses the relevant aspects(s) of the thesaurus, puts "S" (for "please specify !") before it and thus causes the computer to show second highest and afterwards still lower levels of hierarchy; as he goes on, he assigns relevant descriptors by putting "+" before them;
when one or more descriptors have been assigned:		
	"The following descriptors are proposed: (descriptor number) (descriptor number) (descriptor number) a.s.o. " (Two different suggestions are available)	Continues to assign all relevant descriptors by putting "+" before them;
when some 15 ... 20 descriptors have been assigned so that indexing has come to its end:		
		"Please show the complete list of descriptors assigned to this document !": "<u>AUFL</u>"
last no.	Shows complete descriptor list, for printed record.	

The facilities of PRIMAS chiefly result in

- time saved during indexing,
- economics in organization,
- better, particularly more consistent indexing both over long periods of time and with various indexers.

For search purposes, an enquiry has to be entered in the form of a logical connective of descriptors ("and", "or", "and not"), using parentheses whenever appropriate. As results, at first not more than the quantity of relevant documents is given so that the search question might be re-formulated whenever too many, or not enough, documents answer the original question. Thereafter, the searcher gets a list of code numbers of all relevant documents spotted by the computer.

It is therefore expedient to combine PRIMAS with a microfilm system capable of storing full text together with illustrations, graphic representations, "work sheets" covering each document, the content lists and the cover sheets of the periodicals from which articles were indexed, a.s.o. This, no computer could store in any economic way. In cases of such a PRIMAS-microfilm combination, PRIMAS also outputs the addresses leading to the microfilm storage, as an additional service.

Among the many data processing systems, PRIMAS is at present the only one that permits computer-aided input. It is therefore one of our aims to create the conditions in which it can be used as an input system in combination with any existing system: PRIMAS would prepare the input tape for other search programs.

Today already, the current version of PRIMAS offers a system of machine-aided techniques to index or describe documents using sets of descriptor terms characterising the given documents. By offering machine suggestions (and other machine help including various error correcting processes) during indexing, PRIMAS further reduces the intellectual and the "look up" labor of the human document analyst. In fact, PRIMAS incorporates the abilities and choices of a "committee of experts" to help the individual indexer in her, or his task of document description for machine entry.

Moreover, PRIMAS makes it relatively easy to develop, analyze and test the thesaurus and to determine the quality of indexing together with the use of the descriptor terms both for document description and retrieval. PRIMAS also offers

- an "EDIT" function to review and modify existing indexed files and

- a set of automated search/retrieval processes
  to permit the direct search of indexed files.

In addition to these "on-line", interactive machine aids, PRIMAS also includes a set of batch-processing programs which keep track of the above processes and provide the means to perform various statistical analyses for input to file and system managers

The recommended development of PRIMAS constitutes two series or phases of modifications thereto. The first phase incorporates changes suggested by the experimental tests to date. Completion is feasible within 1971 and within current resources. Phase II modifications, however, involve a longer development plan starting in late 1971, and a slight expansion of current resources.

Phase I modifications

The first set of rather uncomplicated improvements includes:

1. Multi-console operation,

2. Expansion of search files (removal of any program-dependent limit),

3. Removal of present 2'000 descriptor limit for indexing,

4. Option for descriptor text wherever descriptor number appears,

5. Reduction of computer operating time for the "SUGGEST" mode,

6. The Bibliographic Data Block
   to be associated with the document accession numbers,

7. Core memory space to be kept lower than 130 kb.

Work on these changes is well under way.

Phase II modifications

Not all users will require the ultimate solution envisioned here, but almost all users will find some of these items most attractive and essential.

1. "Free-text" descriptors

   At the minimum, PRIMAS will have the capability to assign dates as well as names of individuals, organizational entities, and places (or their accepted abbreviations) to documents containing such terms even though such "descriptors" are not among the set of "constructed" descriptors for that field. Given present hardware, such terms would be entered by keyboard

2. Descriptor Qualifiers

In addition to the constructed descriptors, another type of document identification material could be provided: "descriptor qualifiers" refine and further specify the descriptors themselves. For each constructed descriptor, a small set of qualifiers functioning as adjective or adverbial forms could be formulated.

3. Stored Formats

For each constructed descriptor selected by the document indexer/analyst, PRIMAS would offer a "data fill-in" format unique to that descriptor. Other items in such formats could be the request for strongly associated or "linked" descriptors, dates, or names of individuals, organizational entities, places etc.

4. Descriptor Scope Indicators

Certain documents required by some users are already or easily broken down into stipulated document parts such as

title - summary - conclusions a.s.o.;

drawings - diagrams - other graphic portions a.s.o.

Descriptors applying to such documents' parts could be so specified. If the document parts themselves would be identified by sub-accession numbers under the main accession number for the document, search requests could be formulated so as to retrieve the stipulated parts themselves or the entire document or both.

5. Indexing and Analysis for Item or Fact Retrieval

Of course there are occasions when a given descriptor applies to the entire document. In any case from the point of view of the search request, the use of descriptors results in the retrieval of whole documents (or document accession numbers) satisfying the search request. But often it is important to know the specific part of the document (those specific items - or that item - thereof) that triggered the descriptor assignment. This means to specify exactly which textual item or items, phrases or words provoked the descriptor assignment.

The key to item or fact retrieval lies in the development of computer/console programs that enhance and permit the interaction

- of human analysts, textual material and data on one hand

and

- the indexing and analysis apparatus available with PRIMAS.

This technology

- constructed and "free" descriptors, their qualifiers and formats;

- on-line error correcting, feedback, lookup, and "SUGGEST" processes;

- scope indicators and identification of document segments;

needs only the additional capacity for handling full text for controlled display to the human analyst, and the means to activiate freely these PRIMAS procedures, by direct console commands.

This means that both on-line "graphics" consoles and large sized storage are required. Further it is likely that a two screen console - or two consoles, linked - will be needed. It would enable the textual material and scope indicators markers to be handled on one screen while the descriptor selection and subsequent qualifier format display or suggested data occur with the second screen.

LIBRARIANSHIP AND THE USE OF
MACHINE-READABLE BIBLIOGRAPHIC DATA BASES

Irving M. Klempner
Professor, School of Library and Information Science
State University of New York at Albany, U.S.A.

SYNOPSIS

The utopian dream of universal bibliographic control needs to become a reality in the post-industrial era of the 1970's. Machine-readable bibliographic data bases and the ancillary equipment which has recently become available represent significant steps in the realization of the above dream. Present and future incorporation of these data bases into traditional library and information center operations will have profound effects on almost every segment of the information transfer chain, ranging from acquisitions procedures to user feedback. Library educators as well as information service personnel need to adjust their orientation and service concepts to encompass the newly emerging developments.

One of the tasks which I had set for myself is to provide you with some kind of framework, an overview if you like, which would embrace in a broad and general sense the subject to be treated in this paper. I would also like to think that it will be possible for me to place the subject in perspective, i.e., in relationship to past and present bibliographic hopes and current realities.

Permit me to begin with some references to utopian thinking. You see, some of us have at times been characterized as romantics and have been accused of advocating utopian ideas. I believe it only prudent and just to do my part to live up to these characterizations.

Many of you may recall Sir Thomas Moore's work, Utopia, which, as you know, depicts an ideal state, a state in which the curses of poverty, sickness, ignorance and war had been eradicated and where all in society was arranged for the maximum benefit of the individual. Of course, as a label, the term "utopia" was extended to cover earlier and later descriptions of the ideal state. In general, wishing and thinking about utopias was considered to be a relatively harmless exercise. It became the preoccupation of those who sought a convenient vehicle for enumerating and describing the shortcomings of contemporary society, who desired to offer suggestions for vast and revolutionary reforms, and who, having done so, could nevertheless still feel secure, politically and economically. After all, utopians were almost always impractical visionaries who were not to be taken seriously. No one expected utopian writers and thinkers to specify the exact procedures and processes, to provide the detailed flowcharts and programmed instructions by which the ideal state might have been realized. The influence of past utopian projections may be said to have been comprised of interesting intellectual exercises, thought-provoking rather than practical. Certainly most of the individuals who sought to apply utopian ideas to practical realities have often met up with failure, if not worse.

And yet, dreams of utopias, in our particular instance, dreams of bibliographic utopias, die quite hard. The aspirations and efforts in our field leading toward universal bibliography, i.e., the dream of providing rapid and effective access to the totality of man's recorded knowledge, can perhaps be traced to the librarians of Alexandria or the cuneiform tablet-collectors of ancient Babylonia. Scores of individuals, within and outside the library profession, have been projecting and continue to project bibliographic utopias culminating in the total acquisition, organization and servicing of the world's cumulated knowledge resources. The ideas and visionary hopes of the Gesners, Brunets, Otlets and Lafontaines have been supplanted by those of the giant brain advocates, the MEMEX projections of the 1940's and the computer projections of the 1970's.

Writing about the "Fourth Revolution" in a recent issue of the Saturday Review, Isaac Asimov states:[1]

> "The printed word, in a computerized space relay world, will be capable of being transmitted as easily and as widely as the voice...Facsimile newspapers, magazines, and books can be readily available at the press of a button. Perhaps eventually a single world computer will hold in its vitals the library of mankind, any part of which will be available to any man at any time."

Thus is the dream for universal bibliographic control and instant access to documentary resources envisioned.

The idle and visionary bibliographic dreams which prevailed during the more leisurely earlier centuries, and the speculative projections of the more recent bibliographic prophets need to be transformed, as never before, into hard realities today.

We, particularly in the United States, live in what Daniel Bell, Professor of Sociology at Harvard University, describes essentially as the first, "post industrial" nation. While more than 60 percent of the labor force in Asia, Africa and Latin America is engaged in the extractive industries, such as fishing, mining, timber, and agriculture, i.e., preindustrial state occupations, while countries such as the Soviet Union and Japan are engaged primarily in industry and manufacturing, industrial state occupations, the majority of the labor force in the United States is now engaged in services, of which commerce, finance, research, education, and administration form the major and most significant segments.[2] Thus, a pre-industrial society is essentially defined as one based upon raw materials, an industrial society is organized primarily around energy and the use of energy for the production of goods, while a post industrial society "is organized around information and utilization of information in complex systems..."[3]

It is not necessary for me to point out that complex systems, incredibly complex and often in dire need of revision, represent the very essence upon which are based our lives during the era of the 1970's. Information has been defined as the fundamental ingredient essential in the decision-making processes. It is this ingredient, of course, that fundamentally affects the operation and performance of these complex systems. Not only information, but knowledge and wisdom are needed, and will be needed by those responsible for making major decisions and, indeed, by all of us, to devise new systems, to improve existing systems, to make these systems work humanely, efficiently and unobtrusively for the benefit of mankind rather than for its destruction. Neither knowledge nor wisdom can be attained without the essential ingredient of information.

Thus, during the 1970's, more so than in any period of man's recorded history, library and information service personnel have been cast into a pivotal role within our society. As those who have historically geen given the responsibility for storing and maintaining the record of man's experiences and accomplishments, as those who in more recent times have taken on the role of intermediaries in linking that record with current user information needs, librarians and information scientists need to deal not only with a variety of external social, political and technological systems affecting our field, but also with sets of our own internal systems and subsystems. Obviously, it is our bibliographic apparatus, this internal set of systems and subsystems for the acquisition, processing, storage and dissemination of information which provides the key to the recorded discourse of our civilization.

Thus, library and information services personnel are confronted today with the need to make decisions resulting in the acquisition of quality information rather than trash, with the need for selecting and operating economically efficient rather than inefficient systems for information organization, storage and processing, for selecting and making use of desirable rather than undesirable channels or media for information diffusion and dissemination, for disseminating equitably rather than inequitably stored information resources. For information is increasingly being looked upon in our society as a national, indeed, a world economic resource. Broadly defined here as the product of human conceptualization, information is considered also to be an entertainment resource, an aesthetic resource, a moral resource, as well as an economic resource. The culture of poverty in the United States is beginning to be more realistically defined and measured by the degree of lack of information rather than by lack of income or inadequacy of income.

That information, as a human capital input variable, can have profound effects on corporate, regional, national, and, of course, individual well being, has been documented by a new breed of economists, sociologists, and other researchers. Librarians have only recently begun to assign some quantitative and qualitative values to the effects of information on social, political and technological systems prevalent within our society.[4-5]

Let us consider some of the recently developed information system components and see to what extent we have managed to attain the kind of bibliographic utopia sought by our predecessors. The MEMEX machine has not yet been devised; we have successfully linked computer memories within networks, but we have not yet developed the projected single computer with a memory capacity for storing the "library of mankind." Perhaps it is just as well since the wrong bibliographic utopia may have been projected here. In view of our recent experiences with the spectacular and inadequately controlled technological growth in the United States, the question may be raised whether such a mass memory should be developed. It seems to me that it is essential for us to keep in the forefront the thought that our objective is certainly not equipment as an end in itself, but quality bibliographic service for humans. Access to the "library of mankind" may perhaps be attained more economically and humanely through a networks approach rather than via access to a single, massive computer memory. Thus, even were the building of such a computer memory now within the realm of our technology and economic resources, this should not necessarily mean that such a memory ought to be built. Librarians and information scientists, just as any other group, are capable of developing tunnel vision, the kind of vision that often dictates that what can be done, should be done.

While considerable technological progress has been made in creating machine-readable data bases, we have not yet developed the optical scanning devices capable of transforming, at least easily and economically, the cumulated record of man's civilization. The copyright problem is yet to be solved, privacy of data banks is yet to be assured, linguistic problems are still to be defined, software packages are yet to be developed, communications technology is yet to be enhanced.

Notwithstanding the obstacles and difficulties which I have just enumerated, an assessment of the state-of-the-art bearing upon bibliographic utopias leads me to the conclusion that one type of bibliographic utopia, one providing access to the citations which refer to the "library of mankind" is much closer at hand than many of us dare to realize.

Whether a by-product of initial bibliographic publishing activities or direct bibliographic functions of documentation centers or information centers, within the last decade many thousands of machine-readable data bases have become available. These data bases cover almost every conceivable subject and are being packaged and distributed in a variety of formats.[6-7] In a recent issue published by the Information Science and Automation Division of the American Library Association, you will find an announcement to the effect that an ALA Subcommittee on Rules for Cataloging Machine Readable Data Files has been established within the Resources and Technical Services Division, Committee on Descriptive Cataloging.[8] The primary function of the Subcommittee is to isolate requisite points of bibliographic description and to recommend methods of description. The concern of this paper is, of course, not so much the revision and use of the Anglo-American Cataloging Rules for describing machine-readable bibliographic data bases, as it is with the consideration of the use of these data bases within the total bibliographic apparatus that has traditionally been available to us. The fact, however, that such a Subcommittee has been organized is indicative of the growth of the newly emerging information storage medium, the machine-readable data base.

Not only machine-readable information storage media, but also the machines to read, process, and interpret what has been stored will play a vital role in the transformation of traditional bibliographic services. The term machines embraces, of course, new families of computers, transmission channels and services, communication terminals, photocomposition devices and a variety of on-line or off-line equipment for printing facsimile, audio or CRT reproduction and display of information stored in machine-readable form. As far as processes are concerned, the new processes may tend to de-emphasize the significance of the main entry, may bring about changes in ALA descriptive cataloging rules for foreign names, may affect filing and alphabetization rules, may result in changes in classification schedules and notational systems, and may, indeed, result in the introduction of entirely new systems for file organization and processing.

The year 1970 marks the end of the second decade of high speed computers. Costs per computer calculation have improved by factors of 4 to 10 with each new generation of computers. IBM claims that in 1950, one dollar bought the processing of 35,000 computer instructions, in 1967 a dollar bought the processing of 35,000,000 computer instructions. Technological improvements have been attained and are expected in all four main elements of modern computers: logic circuitry, main memory, axiliary memory, and systems organization.

The number of computers in the United States increased from 600 in 1956, to 6,000 in 1960, to 60,000 in 1968, and 25,000 were on order in 1970. A 20 percent annual expansion was forecast for the 1970's. Approximately 100,000 terminals have been produced annually in the United States in the last several years.

If one were to make projections on the basis of the past history of scientific and technological development, there is no question but that we would conclude that additional improvements and innovations in both machines and processes will be made, particularly as they affect computer technology, data base technology and communications technology, all essential elements in the developing of bibliographic utopia of the 1970's. Improvements may be expected in miniaturization, repackaging of information, decentralization of services, programming, marketing, networking and effective feedback from the user as well as feedback from within the internal bibliographic subsystems and broader systems comprising the newly emerging bibliographic utopia.

An analysis of the entire information transfer chain, as outlined and defined by Claude Shannon, leads me to the conclusion that, unquestionably, all of the vital-nodes within that chain will be significantly affected by these projected technological developments. Even if we were to confine ourselves only to the developments which have taken place within the last two decades, were these widely utilized and applied today, they would result in a most drastic transformation of current bibliographic activities and services.

The unfulfilled needs for in-depth information, the unmet needs for equitable distribution of information to all segments of our society, the need for rapid dissemination of information, the need for individualized, personalized, interactive type of information can and will be met in the 1970's regardless of whether librarians and library educators do or do not choose to participate in developing, shaping and refining the bibliographic apparatus capable of supplying these needs. The rapid growth and funding of national documentation centers, clearinghouses, information centers, media centers, professional society bibliographic services, the marketing of bibliographic products and services by industrial organizations attests to the truism that when society is in need of particular service, such a service will be supplied.

Obviously, the library schools and the library school graduates will need to be aware not only of book catalogs and abstracting and indexing services available in book form, but also of data bases available in machine-readable form, whether they be magnetic tape, punched paper tape, disk, cassette, etc.

Illustrative of the kinds of developments that are taking place with respect to storage media for recorded information may be a recent paper presented before the Association of Scientific Information Dissemination Centers meeting in Washington, D. C.[9] The authors, Gloria L. Smith and Joanne Herr of UCRL, note that using film chips, rather than magnetic tape as a basic information storage medium, researchers at UCRL have been able to demonstrate that this particular Mass Storage System "offers a speedy and inexpensive method of information retrieval." For their initial experiment, 4-1/2 years of the Nuclear Science Abstracts data base was used. Significant savings were attained by random searches of the "Chipstore" system as compared to the more sequential searches necessary for the magnetic tape data bases. A recent report from General Dynamics Corporation announced the development of a coding technique, UNIDAR, which makes possible magnetic tape packing densities as high as 33,000 bits per inch, an increase in the order of 90 percent over what is the usual packing capacity.[10] Large scale digital storage, in the range of 10^9 bits and up, is already available through such techniques as laser recording. Other recently announced developments, such as eraseable holographic memories hold promise of extremely high capacities with relatively low space and power requirements.

No doubt, drastic changes in ordering, billing, cataloging and book processing have already been attained through the use of MARC tapes and similar machine-readable records. We can expect that the advent of photo-composition devices and computer controlled preparation of reproduction masters may result in library abandonment of card catalogs in favor of book catalogs and will permit the updating of the catalogs with speed and efficiency.

Significant changes are beginning to occur in the way we use information stores. The reference librarian, if not indeed the user himself, now has or soon will have access either through a terminal or through a local mini-computer, to a linked network system encompassing significant segments of the world's bibliographic resources. Examples from the present which can be projected for the future may be the following: In the latest NASA semi-annual report, we find that the NASA/RECON system of remote consoles for on-line information retrieval which became operational in February 1969 was expanded into a nationwide network of 22 stations linking NASA agency installations throughout the country. The House Committee on Science and Astronautics arranged with NASA to install a RECON console for its use. In a talk given on March 8, 1971 at the American Management Association's Annual Systems Management Conference, Dr. Glenn T. Seaborg, Chairman of the Atomic Energy Commission noted the following:

"At our Oak Ridge National Laboratory we are operating a new computer information retrieval system that provides immediate access to technical literature stored in computer banks some 2,600 miles away. This system is now able to sort through references to about 50,000 technical reports and articles, and within the next few months another 50,000 references will be added. As an experiment in global communications on nuclear research data we recently witnessed the transmission of nuclear information via a new Trans-Atlantic Cable from a computer-based library in California to scientists in Paris. And during the forthcoming International Conference on the Peaceful Uses of Atomic Energy this September we will be transmitting technical data directly via satellite from our Oak Ridge National Laboratory in Tennessee to the conference hall in Geneva, Switzerland." 12

The study of the many technological developments in the information science field and the study of the more humanistically-oriented field of librarianship leads me to the conclusion that none of us need to cling to the printed codex or manuscript to find reason and morality in the 1970's. It seems to me that all of us, librarians and information scientists, must learn to live with the newly developed technological capacity, and live up to the new responsibilities it brings, not merely because it is here, but because developed and applied wisely it will help us achieve the most human goals.

1. Asimov, Isaac. "The Fourth Revolution." Saturday Review, October 24, 1970. p.19
2. U.S. Congress. House. Committee on Science and Astronautics. The Management of Information and Knowledge. Washington, D. C., Govt. Print. Off., 1970. p.13-15.
3. ibid. p.14.
4. Olsen, Harold A. The Economics of Information: Bibliography and Commentary on the Literature. Washington, D.C., ERIC Clearinghouse on Library and Information Sciences, January 1971.
5. Klempner, Irving M. "Information as an Economic Resource." In Diffusion of Abstracting and Indexing Services for Government-Sponsored Research. Metuchen, N. J., Scarecrow Press, 1968. p. 32-49.
6. Carroll, Kenneth D. Survey of Scientific-Technical Tape Services. New York, American Institute of Physics in conjunction with the American Society for Information Science, September 1970. (AIP ID 70-3; ASIS SIG/SDI 2)
7. Cohan, Leonard, ed. Directory of Computerized Information in Science & Technology. New York, Science Associates/International, Inc., 1968-69.
8. "Cataloging Machine Readable Data Files." JOLA Technical Communications, v. 2, no. 3, March 1971, p. 1.

9. Smith, Gloria L. and J. Joanne Herr. Searching the "Nuclear Science Abstracts" Data Base by Use of the Berkeley Mass Storage System. Berkeley, Calif., Lawrence Radiation Laboratory, University of California, Berkeley, February 1971. (UCRL-20286).

10. Piasta, Frank. "33,000 Bit/in Possible through Coding Method." Computer World, April 29, 1970, p. 4.

11. National Aeronautics and Space Administration. Semiannual Report, July-December, 1969. Washington, D.C., U.S. Govt. Print. Off., 1971. p. 157.

12. Seaborg, Glenn T. Technology, Information and Computers. Paper presented at the American Management Association's 17th Annual Systems Management Conference, New York, March 8, 1971.

Session Ten - Discussions

PROCESSING FOR AUTOMATION - II

Chairman: Prof. Th. P. Loosjes (Netherlands)

MR. J.H. D'OLIER (France): The basic principles of the P.A.S.C.A.L. (Programme Applique a la Selection et a la Compilation Automatiques de la Litterature) system can be summarized as follows:

First, our customers require information more and more adapted to their specific needs, so we have to offer them more and more varied and specialized services.

Second, to provide such services, it is necessary either to create them independently or, after suitable adaptation, to use services gleaned all over the world. However, the cost involved would be too much for most of the French users, even for medium-sized industrial firms.

Third, we know that several operations carried out on original documents are almost the same for a range of end products and services.

Fourth, an information center in science and technology which would cover all the diversified and precise questions of our customers does not exist. Therefore, only a cooperative system can be worked efficiently and economically.

Such a system must be studied and started from the beginning on a cooperative basis. To answer the requirementa stated above, it is necessary to study with the participants the operations of documentation processing so as to come to precise and detailed agreements.

What are the main operations? There are intellectual ones, such as abstracting and indexing, and technical ones, like transcribing and transforming bibliographic descriptions according to international rules and acceptable for computer processing.

This has to include all the services and all the products which the system will offer. The worksheet form has been thoroughly studied. Many corrections were necessary but we came to the conclusion that the cost of corrections is lower than the costs which would result from duplication of some of the operations.

The most important tool is the thesaurus. We are constructing a French thesaurus using as a basis many existing thesauri - American and others. We are also constructing a bi- and multi-lingual thesaurus in consultation with documentation centers in other countries. However, we had to start work without a thesaurus, adding new terms as we go and we have a computer program which simultaneously constructs a thesaurus and performs selective dissemination of information with the available vocabulary which has been entered since the beginning of the operation.

From this in-put, the system produces a bibliographic bulletin, with author indexes and a three level subject index. These indexes do not use the same thesaurus as the classification scheme but are complementary to it. The computer also produces cummulative annual author and subject indexes in normal typography. This month we shall start with SDI, but profiles will have to be in the vocabulary of the system. In medicine, pharmacology, arts and sciences - fields in which we have thesauri - according to these thesauri; and in fields where we don't yet have them - in our own vocabulary. In fields in which we don't have a thesaurus we prepare the questions ourselves since we know our vocabulary, although it is not yet published and the users do not. The vocabulary is specific to the different fields of science, but we are also preparing a macrothesaurus to bridge the fields.

Next year we hope to offer SDI services in about 30 disciplines in science and technology, thanks to close cooperation with more than ten documentation centers, both in France and abroad.

We also intend to offer magnetic tapes and to begin retrospective searching. But so far we have only one year's accumulation for retrospective searches. In 1971 we shall have stored about 100,000 documents with their keywords, authors, titles, and bibliographical references. Next year we shall store 300,000 items, and later we shall reach 500,000 per annum. This is too much for retrospective searching. After the first year it is less expensive to maintain a photomemory for abstracts and bibliographical references, and to keep on magnetic tape only keywords, authors and document numbers.

An automatic connection will be arranged between the computer and the photomemory. Therefore, from the beginning of next year our Bulletin Signaletique will be published as hard copy and on microfiche. With an automatic system the search is much easier with microfiche than with a hard copy.

DR. W.E. BATTEN (UK): I offer you two new pieces of terminology to embrace the curious topic of profile improvement: "profile cosmetics"; or, if you are more given to puns, "profilactics". Either of them will cause the thesaurus makers the usual headaches.

The work described in our paper was done by Dr. Veal and his two colleagues, and my only personal contribution was to say occasionally, you cannot have any more money, or, will you please run a little faster.

The present line of study will probably continue for many years, and it is much too early yet to attach operational significance to the present figures. For the present, we do seem to be moving in a vaguely hopeful direction, but violent corrections of course may be expected several times before we reach our objective.

One of our earliest observations as an operational research center was that a user's initial profile could usually be improved upon. There were two aspects to this; first, users of computerized services were not accustomed to being highly explicit. They had been used to having search assistants who had a modicum of scientific know-how and common sense. These assistants could move from the generic to the specific without any programmed instructions, and great reliance is placed on this common sense and on whatever familiarity with the subject the researcher has.

Secondly, the center itself was none too successful at first in translating the natural question into a search specification even when the interfacial adviser was himself or herself reasonably familiar with the field of inquiry.

We decided to leave the creation of explicitness to our service department because it is largely a matter of inducing in the inquirer a certain state of mind. On the other hand analysis seems to indicate that the user is not automatically aware of the richness of the vocabulary of his speciality. He may be still less aware that there may be valuable points of access that have no semantic connection with his question. Incidentally, Science Citation Index could be regarded as an extreme result of this observation, although it came into existence for other valid reasons.

Our opening thesis was as follows: examining the out-put from an initial profile we find both unsatisfactory recall and unsatisfactory precision. This resulted in a series of research projects conducted with the cooperation of the user. We launched a series of experiments to discover what, if any, improvement could be obtained by an analysis of the retrieved items and the methodology was briefly this:

The user was given a period of ordinary SDI service to establish an initial performance rating for his profile. We then put him on to a period of searching the same profile against a fixed file of references. Then came a further period of SDI service, using the profile that had resulted from the "iterated" use of the fixed file. This gave a second SDI performance rating for the same inquiry and hence enabled improvement, if any, to be observed. Please note I decline to say "measured" - I will stick with observed.

Now, the process of searching the fixed file was iterative. By intellectual and statistical examination of the output, it was possible to make changes in the component terms and sometimes in the logic of the profile. The amended profile was then searched against the fixed file. This resulted in yet further relevant output which was again examined and the profile again amended and this was continued until there was no further addition to the relative output. Usually, three iterations were sufficient. The perfected profile was used for the second comparative period of SDI service.

A further effect was studied at the same time. That is the relative performance of using titles only as the search basis for the same question. At the recently concluded Cranfield Conference in July 1971 it was possible to give more detailed analysis than at the time of preparing my paper. These results will be available in print soon.

Now, I will say a little more about the profile-improvement study. A campaign of term analysis was conducted. For this purpose we recognized four types of terms:

1. Specific terms, which appear mainly in relevant items.
2. Non-specific but useful terms, which appear in fairly equal portions in both relevant and irrelevant items.
3. Redundant terms, which appear either very infrequently or always in the presence of other more frequent terms of type 1 or 2.
4. Detrimental terms, which appear mainly in irrelevant items.

When a user improved his profile by hand, he was unconsciously classifying the potential terms in accordance with these four headings. It ought therefore to be possible to mechanize the process of profile refinement if only we could find valid parameters to enable term reference to be expressed. We felt that the terms must have something to do with the frequency with which they cropped up in the out-put. We adopted the concept of specificity value, which we define as being the number of relevant documents in which the term occurs, divided by the total number of documents in which it occurs.

This is a ratio which lies between 1 and 0. We took all the terms which had assisted in the production of every relevant document, compiled specificity values and listed the terms in order of specificity value and then according to the later profile, according to the spread of the specificity list. We recommend to the user to make use of all terms lying above a certain specificity level. When the Cranfield Conference papers come out you can see the tabulated results. The qualitative and the quantitative results of the analysis show:

1. Automatic analysis was useful since extra relevant items were retrieved from

the fixed files, sometimes in large numbers. This in itself was sufficient to convince users that their search profiles were missing interesting information and they should look at them more closely.

2. This approach avoids the necessity of writing a statement of interest. This may be of some use to people who find it difficult to make a statement of interest. The alternative is for the user to say that he can't find the words to describe his interest, but does know that the following three references are relevant. You can then start with these references and pick the terms out of them.

Automatic analysis of relevant items produces a list of terms in which specific terms appear first, followed by non-specific terms, followed by the more detrimental terms. The method is still very crude since at present only single terms are taken into account. It, therefore, produces the best results for single concept interests and progressively worse results for two-concept and multiple-concept interests.

The specificity may sometimes suggest quite an unusual search semantically. E.g., in a search originally written as a two-parameter search, lead terms were linked with toxicity terms. The specificity list suggests that the statement of interest should really have been "air pollution by lead and mercury compounds" and that better results would be obtained by searching for city names and terms concerned with traffic which are specific to the topic. The original recall was 74%. The improved recall was 95%. That's for CA Condensates 1. On CA Condensates 2, the improvement was from 88% to 95%. The improvement was obtained by advisedly adding terms from the specificity list. I am bound to tell you that when we went by the specificity list only, the initial recall was 85% but the improved recall was only 70%, and on CAC-2 even worse, the original recall being 90% and the improved only 33%. So you can have too much of a good thing.

The procedure used so far aids in the location and ordering of the more specific single terms pertinent to a given topic. If all fields of interest could be described in single terms, this procedure, with some refinements, such as choice of the "best word fragments", would suffice for constructing a fairly adequate profile or refining an existing one. However, the majority of queries contain at least two concepts linked by AND logic and although some single terms may be useful, the bulk of material is retrieved by term pairs. Therefore, we are now applying the same technique to an examination of the frequency and incidence of term pairs. To date, we have carried out only one manual experiment and the results are interesting. The original profile for this particular test topic contained 40 terms and achieved 24% precision and 69% recall with an output level of 34 per run; this profile would cost £100 per year at the present pricing structure. With single terms from the specificity list, this level of recall could be achieved at a cost of about £90 per year. With a mixture of single specific terms and term-pairs, it could be achieved at a cost of about £80 per year.

I don't think this will lead us to the ultimate perfection in construction. What it will show, is the shape of the performance curve according to the discriminatory power put at the user's disposal. We shall be able to plot that against a likely price, so that a user who says, "I can pay £50," won't get the best curve but the best value for his money.

DR. E. HOFFMANN (Israel): As chemistry develops, the chemical code must develop. The structural formula of today is becoming inadequate to represent meaningfully all chemical information about a certain compound. There are all kinds of secondary aids with which the chemist tries to get around this, but somebody should think about a different, more precise and up-to-date representation of chemical knowledge than the present use of the chemical formula.

The enormous amount of information in chemical structure produced today forces us towards mechanization.

If you squeeze the structural formula into the computer by ultranumeric means, you are liable to have an information loss; transfer from one code to another usually involves an information loss, whereas what you want is an information gain.

You thus face a double dilemma. In order to enable your colleagues to reach all the information, you have to go to the computer, but by doing this you incur a rather substantial information loss, which, considering the enormous development of the field, may still grow in coming years.

In my paper I have reviewed the three categories of codes which are in use today, and their various capabilities and uses. Many chemical codes can be put into these three categories. Since chemical information systems were not developed at universities but in industrial mission-oriented organizations, chemical codes were adopted for specific purposes. They were not meant to recover the total of chemical information, but to do a specific job which they do very efficiently. Big chemical companies in the US and the UK have their own information systems which use codes developed internally.

This fact leads us to the problem of compatibility. There is no easy way to transfer one code to another mechanically without a conversion program. The problem is the conversion into a topological code or from the topological code, because the topological code is the code with the most information. It is also the most expensive code to operate. The vast store of Chemical Abstracts, archival in nature, is stored in a topological system.

There is a great demand to make this store available in such a way that it can be searched economically. This is impossible as long as the storage is in a topological system. There are efforts to transform this system into one which would be cheaper to operate, with an inevitable information loss. There is a possibility of using complementary coding where information is kept in different codes.

The other type of conversion is into the topological code. You encode your formulas compactly and store them with relatively little expense. If you want to get complete information, which you do not have because you did not use the topological code, you have to transfer your store into a topological system.

There is a third way of intermingling the various codes economically. You could - and this is being done - encode the topological code. The encoding procedure is very simple and does not require highly qualified manpower. Once you encode into a topological system, you can transfer it into another code better adapted for compact storage, and from that storage you can query economically.

MR. H. SCHUR (UK): *To Mr. d'Olier.* What are the problems of using French for indexing? Do you use in the keyword index a mixture of English, French and other languages? In other words, do you translate first into French and have French keywords only or do you have a mixed language keyword base?

MR. D'OLIER: *To Mr. Schur.* We have a French thesaurus and French keywords, but our thesaurus and our vocabulary are closely related to the English vocabulary and to English keywords. For a question with non-French keywords, we translate with the help of a multi-lingual dictionary. We are, however, planning to have English and German keywords in many disciplines. For instance, in medicine, we have a tri-lingual thesaurus, French, English and German. In arts and science, our thesaurus is developed in close cooperation with other science centers in the world. In meteorology we are building three and five language thesauri. In Luxembourg there is such a program for meteorology. If you ask a question in German you can have the reply in German although the system works in English.

QUESTION: To Mr. d'Olier. As you know, ESRO (European Space Research Organization) has a very important information center in Paris. It is working with English tapes and intends to increase its services. Has the CNRS studied the possible over-lapping of these two services both located in the same place?

MR. D'OLIER: Space science is not one of the fields we dealt with in the beginning because there exist several documentation centers in the field and there is a lot of information. We are dealing with some aspects of space science but with a different orientation from ESRO. Our efforts are especially oriented toward our own industry and research in astronomy. We do not intend to duplicate existing services, and in space science English is predominant so in this field it is less necessary to have a French documentation system. We are developing a French system with French keywords because documentation must be aggressive and to command attention must talk to the user in his mother tongue.

ESRO works entirely in English. If it supplies information suitable for French users we do not intend to duplicate it, but if some users require another presentation, maybe we will cooperate with ESRO. There is another field in which French may not be necessary - atomic energy. Atomic energy today has wide applications. Some documentation is desirable in French for its applications in medicine, plastics, chemistry, meteorology, etc.

MRS. E. KATZ (Israel): To Mr. d'Olier. You state in your paper that you use 9,000 journals as your source material. What is the percentage of Chinese language journals that you analyze and in what fields?

MR. D'OLIER: To Mrs. Katz. Zero. We haven't used any Chinese publications since the beginning of the cultural revolution. But today we think that perhaps the Chinese language will have to be scanned.

PROF. LOOSJES: To Dr. Batten. Do you treat documents with partition when partition of certain documents makes the analysis more specific?

DR. BATTEN: To Prof. Loosjes. That is very much a matter of the content of the document. We would not presume to divide an item which in the author's view was not heterogenous. But, if it were the abstract of a symposium where there were several related contributions all following one general broad theme, you needn't avoid partition.

MISS M. PARK (US): To Dr. Hoffmann. On page 14 of your paper you cited a figure of 3 million, and I wondered if that referred to the number of unique connectivity tables in that file, or the number of unique compounds. Our number is about half that.

DR. HOFFMANN: To Miss Park. I meant the number of compounds.

MR. D'OLIER: To Dr. Hoffmann. In France we are developing a compact topological coding system which has the advantage of keeping all the information in the topological code compactly enough to be handled by the computer. Do you think it would be possible to have some cooperation between your work and our system?

DR. HOFFMANN: To Mr. d'Olier. I must confess my ignorance of your system. Of course there are other topological systems which you can produce from line notations. I have in mind ICI's Rossbow system. I think it is worthwhile to look into this.

QUESTION: To Dr. Hoffmann. In France we are developing auto-electronical methods which make it possible to solve some of these problems. A balanced system between computers and auto-electronic systems could solve some chemical compound problems.

DR. HOFFMANN: That is quite true. The question is at what price. The methods I have seen in the US are still much too expensive for ordinary users. It will take a long time until the facilities are available and the price has come down sufficiently to consider operating these methods.

DR. BATTEN: There is one other subject where almost the whole of its compact language could be properly described as a code and that is music. I have often hoped that somebody would make an economic study of chemical topological codes and musical codes. There is knowledge in both that might be mutually borrowed.

MRS. L. VILENTCHUK (Chairman, ISLIC): Dear Colleagues, Guests and Israelis: with this session the Conference comes to a close. I thank you for coming and for making this Conference a success, by chairing the sessions with patience and humour, by evaluating the papers before the Conference, by participating in the discussions from the floor, by giving us, the Israelis, the opportunity to cement new contacts and friendships. I hope you have enjoyed the Conference and your stay in our country, and I have only to add, come again soon! Thank you.

Open Meeting of FID/II

SUMMARY*

Chairman: Mr. C. Keren (Israel)

In continuation of the conference an open meeting was arranged to give the representatives from the FID Study Committee "Information for Industry" (FID/II) an opportunity to present their views and experiences as to how information is communicated to users in industry.

Mr. Kjeld Klintøe (Chairman, FID/II) first presented the underlying idea of FID/II emphasizing the link between economic growth and the systematic application of knowledge. A description was given of the fields of interest of FID/II and of the various projects.

Mrs. V. Vince (UNIDO, Vienna) described the activities of the Industrial Services and Institutions Division. This section of UNIDO assists developing countries in how to organize industrial information facilities and how to get access to the information stored in the industrialized countries.

Mr. D.G. Kingwill (CSIR, South Africa) outlined the organization of the Council for Scientific and Industrial Research and explained how the extension services functioned, especially those serving medium and small-sized industry.

Mr. A. Disch (SNI, Norway) mentioned some of the activities of SNI which is a center for technical information and documentation. The center publishes abstract journals for the practical engineers in industry and has developed a computerized retrieval system, POLYDOC.

Mr. F.G. Halang (National Research Council, Canada) surveyed the organization of the Canadian information services working for secondary manufacturing industry. As Mr. J.E. Brown had previously talked about the National Science Library, Mr. Halang concentrated on the Technical Information Service with its field services for industry and stressed the importance of the human approach in the transfer of information.

Mr. Kjeld Klintøe (DTO, Denmark) described the idea behind the Danish Technical Information Service, an organization for marketing knowledge, whose fundamental activity is the field liaison service. The special importance of interviewing managers in industry was accentuated. Mr Klintøe also described the SDI and referral activities and mentioned the shift in interest from productivity methods over product development to long-range planning.

Mr. F. Liebesny (ASLIB, United Kingdom) saw the situation in the UK from three viewpoints: subject oriented dissemination of information carried out by the research associations, regional covering of industry's needs through the industrial liaison officers, and field oriented information centers such as ASLIB.

Editors' note: Thanks are due to Mr. K. Klintøe for putting this summary at our disposal.

Dr. A. van Loen (NOBIN, Netherlands) described some of the problems facing the newly established NOBIN, which is an advisory body to the government in information matters.

Mr. C. Keren (COSTI, Israel) briefly mentioned the liaison service serving industry emphasizing the viewpoint that information is sold like any other product.

Dr. M. Cremer (Inst. für Dokumentationswesen, German Fed. Rep.) emphasized the growing interest of industry in an integrated information system comprising law, regulations and other data.

During the discussion following the presentations, *Prof. P.V. Kaula* described the activities in India in the field of dissemination of information to industry.

As it was felt that the presentations at the open meeting of FID/II contained much valuable material which deserves to be made more widely available, the Proceedings will be published as a special publication in the ISLIC series "Contributions to Information Science," publication No. 6, and can be ordered from the ISLIC Secretariat, P.O.B. 20125, Tel Aviv, Israel.

list of participants

AUSTRIA

BLANC, Dr. Hofrat W.
Vorstandsvorsitzender,
Rechenzentrum Graz,
A-8010 Graz, Steyrergasse 17

RAIMANN, Dr. Herbert
Director, Rechenzentrum Graz,
A-8010 Graz, Steyrergasse 17

SIMMLER, Dr. Otto
Direktor, Bundeskanzleramt Wien,
A-1010, Wien, Herrengasse 23

BELGIUM

COCKX, Dr. A.
Director, Royal Library of
Belgium - CNDST, 4, Boulevard
de l'Empereur, Brussels

CANADA

*AMEY, Dr. G.X.
Head, Information System Development,
Defence Research Board, Ottawa 4,
Ontario

*BROWN, Dr. Jack E.
Director, National Science Library,
100 Sussex Drive, Ottawa 7, Ontario

HALANG, Frederic G.
Secretary, Advisory Board on Scientific
and Technological Information, National
Research Council of Canada, Ottawa 7,
Ontario

DENMARK

FINK, Dan
Head, The Danish National Center for
Building Documentation, Copenhagen,
DK 1055

HVIDEGAARD, Torben
Civil Engineer, Denmark Technical
Information Service (DTO), Ørnevej 30,
2400 Copenhagen NV

* Speaker

DENMARK, con'd.

KLINTØE, Kjeld
Director, Denmark Technical Information
Service (DTO), Ørnevej 30, 2400 Copen-
hagen NV

*SKOV, Helge J.
Danmarks Tekniske Bibliotek,
Dokumentationsafdelingen Anker,
Engelunds Vej 1, 2800 Lyngby

FRANCE

ARIES, Philipe
Director, Center of Documentation,
Institut Français de Recherches (IFAC),
6, Rue du General Clergerie, Paris 16e

CESTAC, Mrs. Françoise
U.F.O.D., 16 Rue Jules Claretie,
Paris 16e

COUSTURE, Miss Micheline
22, Rue de Vouillé, Paris 15e

de LACLEMANDIERE, Jean
U.F.O.D., 16, Rue Jules Claretie,
Paris 16e

LEVY, Miss Denise J.
Groupe des Laboratoires du CNRS de
Vitry-Thiais, 2 à 8 rue H. Dunant,
94-Thiais

MARAL, Mrs. Francqise
U.F.O.D., 16, Rue Jules Claretie,
Paris 16e

*d'OLIER, J.H.
Directeur-adjoint, Centre de Documen-
tation du CNRS, 15, Quai Anatole
France, Paris 7e

GERMANY

*ARNTZ, Prof. Dr. Helmut
Vice-President, FID, 534 Bad Honnef/
Rhein, Burg Arntz

GERMANY, con'd.

BUDINGER, Dr. Peter
Wissenschaftlicher Dokumentar, Zentralstelle fuer Maschinelle Dokumentation, 6 Frankfurt, Herriotstrasse 5

CREMER, Dr. Martin
Director, Institut fuer Dokumentationswesen, 6 Frankfurt(Main)/ Niederrad, Herriotstrasse

HEINZE, Helmut
All Delphi, 2 Hamburg 1, Postfach 992, Steindamm 94

HENRICHS, Dr. Norbert
Associate Director, Department of Information Science, Philosophical Institute, University of Dusseldorf, Alte Landstrasse 28, Dusseldorf

HILTMANN, Prof. Dr. Hildegard
Lehrstuhl fuer Angewandte Psychologie Universitat, 78 Freiburg 1. BR. Peterstrasse 1

*KRALLMAN, Prof. Dr. Dieter
Privat Dozent, University of Bonn, Adenauerallee 98a, 53 Bonn

NIEDERMEYR, W.J.
Division Leader, Information Centre, Nuclear Research, Kernfaschugszentrum, Karlsruhe

*REICHARDT, Dr. Guenther
Bibliotheksleiter, Zentralbibliothek der Kernforschungsanlage Juelich GmbH, D 5170, Postfach 365

*SCHUTZSACK, Dr. Udo
German Manager, International Food Information Service (IFIS), Institut fuer Dokumentationswesen, 6 Frankfurt (Main)/Niederrad, Herriotstrasse

STEINBACH, Dr. W.
Stiftung Volkswagenwerk, 3000 Hanover 26, Postfach 260509, Kastanienallee 35

STEINERT, D.L.
Head of Information Science, German Foundation for Developing Countries, Central Documentation Branch, 53 Bonn, Endenicher Strasse 41

GERMANY, con'd.

*THIELE, Georg
Leiter der Arbeitsstelle fuer Reprotechnik, Institut fuer Dokumentation, 6 Frankfurt/Main 1, Westendstrasse 19

VOSS-ECKERMANN, Miss Helga
Research Assistant, Research Institute for Legal Documentation, 31 Senckenberganlage, Frankfurt/Main

*WAGNER, Prof. Dr. med. Gustav
Director, Institut fuer Dokumentation, Information und Statistik, 69 Heidelberg 1, Berliner Strasse 27

ZIMMERMANN, Gerhard
Praesidium der Deutschen Gesellschaft fuer Dokumentation, D6 Frankfurt 1, Westendstrasse 19

*ZUCHEL, Horst
Programmier-Abteilung, Zentralstelle fuer Maschinelle Dokumentation, 6 Frankfurt(Main)/Niederrad, Herriotstrasse

INDIA

*KAULA, Prof. Prithvi N.
Head, Department of Library Sciences, Banaras Hindu University, Varanasi-5

ISRAEL

ACHITOV, Mrs. Miriam
Vulcani Institute for Agricultural Research, POB 6, Beit Dagan

ADLER, Elhanan
Head, Department of Cataloguing and Classification, The Library, University of Haifa, POB 5345, Haifa

AGMON, Y.
Elta Electronics Industries Ltd., POB 330, Ashdod

ALLOSS, Eliyahu
Head, Physics and Chemistry Library, Hebrew University, Jerusalem

ALONI, Mrs. Michaela
Armaments Development Agency, Ministry of Defence, POB 962, Haifa

ISRAEL, con'd.

AMIEL, Mrs. Esther
Documentation and Information Center, Office of the Chief Scientist, Ministry of Defence, Tel Aviv

ARNON, Mrs. Louise
64 Herzl Avenue, Jerusalem

ARNON, Mrs. Tamar
Documentation and Information Center, Office of the Chief Scientist, Ministry of Defence, Tel Aviv

ARON, Mrs. Ellen
Programmer, National Center for Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

ASAF, Mrs. Mira
Documentation and Information Center, Office of the Chief Scientist, Ministry of Defence, Tel Aviv

AVGAR, Dan
Israel Atomic Energy Commission

AVIVI, Dr. Ehud
Nuclear Research Center - Negev, POB 9001, Beer Sheva

AVNI, Hanan
Director, Documentation and Information Center, Office of the Chief Scientist, Ministry of Defence, Tel Aviv

AVRAHAM, Rami
Documentation and Information Center, Office of the Chief Scientist, Ministry of Defence, Tel Aviv

AVRIEL, Mrs. Deborah
Jewish National and University Library, Julius Jarcho Medical Library, POB 503, Jerusalem

BARAK, Nachman
Head of Training, Ministry of Defence, The Kirya, Tel Aviv

BAR GAD, Mrs. Nurit
The Library, University of the Negev, POB 2053, Beer Sheva

ISRAEL, con'd.

BARUCH, Mrs. Ruth
Department of Acquisitions, The Library, University of the Negev, POB 2053, Beer Sheva

BEGED-DOV, Prof. Aron
Department of Applied Mathematics, Weizmann Institute of Science, POB 26, Rehovot

BEN-AMI, Mrs. Sarah
Ministry of Defence, POB 962, Haifa

BEN CHAIM, David
Head of Classification Division, University of the Negev, POB 2053, Beer Sheva

BEN YOSEF, Miss Hadara
Librarian, Armament Development Agency, Ministry of Defence, POB 962, Haifa

BEN-ZION, Col. Elka
Librarian, MPO 1002, Israel Defence Forces

BIRAN, Dov
Elta Electronics Industries Ltd., POB 330, Ashdod

*BLOCH, Uri
Armament Developing Agency, Ministry of Defence, POB 962, Haifa

BOEHM, Mrs. Alina
Chief Librarian, School of Education, Hebrew University, Jerusalem

BRACHA, Dr. Peretz
Makhteshim, POB 60, Beer Sheva

BURDMAN, Miss Edith
Assistant Librarian, National Center for Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

BZISKY, Mrs. Lea
Documentation and Information Center, Office of the Chief Scientist, Ministry of Defence, Tel Aviv

CAFRI, Mrs. Donia
Librarian, The Standards Institute of Israel (SII), Ramat Aviv

ISRAEL, con'd.

CANTONI, Dr. A.
Electrochemical Industries (Frutarom) Ltd., POB 1929, Haifa

CANTONI, Mrs. Ahuva
Chief Librarian, Library of the Department of Agricultural Engineering, The Technion, Haifa

CASSUTO SALZMANN, Mrs. Milka
Head of Library, Department of Antiquities, Ministry of Education and Culture, Jerusalem

CHACHAM, Mrs. Mazal
Ministry of Communication, Jerusalem

CHARNY, Mrs. Nadine
Librarian, National and University Institute of Agricultural Research, POB 12, Rehovot

CHWOLLES, Mrs. Helene
Librarian, Faculty of Civil Engineering, The Technion, Haifa

COHEN, Mrs. Ana
Programmer, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

*COHEN, Israel A.
Programmer, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

COHEN, Miss Pauline
Volcani Institute of Agricultural Research, POB 6, Bet Dagan

COHEN, Mrs. Tirtsa
Cataloger and Classifier, Wix Library, Weizmann Institute of Science, POB 26, Rehovot

CORET, Mrs. Alice
Nuclear Research Center - Negev, POB 9001, Beer Sheva

DAGAN, Mrs. Afrah
University of the Negev, POB 2053, Beer Sheva

DIKOVSKY, Shlomo
Ministry of Commerce and Industry, Jerusalem

ISRAEL, con'd.

DJERASSI, Mrs. Mathilda
Head, Journal Division, Central Library, Tel Aviv University, Ramat Aviv

*DORON, Meir
Production Division, Keter Publishing House Ltd., Israel Program for Scientific Translation, POB 7145, Jerusalem

DROR, Mrs. Lea
5 Kortzak Street, Holon

EHRLICH, Mrs. Yocheved
Documentation and Information Center, Office of the Chief Scientist, Ministry of Defence, Tel Aviv

EHRLICH-DEVRIES, Mrs. Esther
Librarian, Mineral Engineering Department Library, The Technion, Haifa

EISENBERG, Mrs. Lea
Assistant Editor, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

ELATH, Matityahu
University of the Negev, POB 2053 Beer Sheva

ELPERN, Leon
Head of Library, Truman Center, Hebrew University, Jerusalem

ETHAN, Mrs. Bracha
Information Scientist, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

EVEN TSUR, Mrs. Dalia
The National Cash Register Company (NCR), 1 Balfour Street, Tel Aviv

FARBER, Mrs. Miriam
51 Aharonson Street, Ramat Gan

FARHI, Mrs. Edna
Information Scientist, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

ISRAEL, con'd.

FELDMAN, Miss Yehudith
Central Library, Israel Air Force, MPO 2704

FINKELSTEIN, Meir
Engineer, Institute of Current Technology (MTM), POB 16108, Tel Aviv

FONIA, Miss Alisa
Librarian, Israel Defence Forces, MPO 1068

FRANK, Miss Mathilda
Israel Defence Forces, MPO 2475

FREIDIN, Aharon
Israel Aircraft Industries, Lod Airport

FRIMET, Mrs. Sidonia
Librarian, Faculty of Agricultural Engineering, The Technion, Haifa

GASSMAN, Uri
Medical Librarian, The Library, Beilinson Hospital, POB 85, Petah Tikva

GDALEWICZ, Benedyct
Director, Information International Ltd., 25 Shlomzion Hamalka Street, Tel Aviv

GILBERT-GIL, Beno
Information Officer, Soreq Nuclear Research Center, Yavne

GILLAM, Mrs. Miryam
Acting Head, The Library, Association of Engineers and Architects in Israel, 200 Dizengoff Road, Tel Aviv

GLASSMAN, Mrs. Yaffa
Chief Librarian, Negev Institute for Arid Zones Research, POB 1025, Beer Sheva

GOLDBERG, Mrs. Margalit
Librarian, Metals Library, The Technion, Haifa

GOLDENBERG, Mrs. Nurit
Programmer, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

GOLLOP, Mrs. Erna
Dead Sea Works, Ltd., Potash House, POB 75, Beer Sheva

ISRAEL, con'd.

GOREN, Mrs. Rivka
Head of Library, Israel Petroleum Institute, 26 University Street, Ramat Aviv

GROSS, Mrs. Barbara
Acquisitions Department, Soreq Nuclear Research Center, Yavne

GROSS, Ernest
Information Technologist, Armament Development Agency, Ministry of Defence, POB 962, Haifa

GROSS, Prof. Jacob
Chief Scientist, Ministry of Commerce and Industry, Jerusalem

HARTMAN, Mrs. Marietta
Documentation and Information Center, Office of the Chief Scientist, Ministry of Defence, Tel Aviv

HARRAN, Mrs. Tina
Head, Library, Machteshim, POB 60, Beer Sheva

HEMMERDINGER, David W.
Water Planning for Israel (Tahal), 54 Ibn Gvirol Street, Tel Aviv

*HOFFMANN, Dr. Eliahu
Head, Division of Documentation, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

HORAK, Miss Miriam
Technical Secretary, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

*ISAACSON, Sheldon
Manager, Center for Information and Applied Technology, Koor Industries Ltd., Tel Aviv

JAGUR, Mrs. Orna
Librarian, Israel Institute for Biological Research, POB 19, Nes Zionna

ISRAEL, con'd.

KADMON, Gedaliah
Ministry of Defence, Haifa

KAHANA, Mrs. Ora
Head of Library, Computer and Mechanized Recording Center (Mamram), Ministry of Defence Tel Aviv

KALB, Mrs. Fay
Librarian, Department of Industrial and Management Engineering, The Technion, Haifa

KAMIN, Mrs. Rosalinda
Librarian, National Council for Research and Development, Prime Minister's Office, Building 3, The Kirya, Jerusalem

KAPLAN, Mrs. Eva
Librarian, Weizmann Institute of Science, POB 26, Rehovot

KARLIBACH, Mrs. Sima
Central Library, Israel Air Force, MPO 2704

KARNI, Mrs. Miriam
Chief Librarian, Road Safety Center, 27 Adam HaCohen Street, Haifa

KATZ, Mrs. Edith
Editor, Weekly Bulletin of the Technion, Haifa

KATZ, Mrs. Rosa
Ministry of Defence, Haifa

*KEREN, Carl
Director, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

KESSEL, Mrs. Shoshana
Nuclear Research Center - Negev, POB 9001, Beer Sheva

KICHALES, Mrs. Simona
Meteorological Institute, POB 25, Beit Dagan

KLAUS, Uri
Librarian, Israel Defence Forces MPO 1002

ISRAEL, con'd.

KLEIMAN, Mrs. Hava
Jewish National and University Library, Hebrew University, Jerusalem

KOLIN, Mrs. Miriam
Abic, Chemical and Pharmaceutical Industries, POB 2077, Ramat Gan

KOPPEL, Benedictus
Director, Library and Scientific Documentation Unit, Israel Institute for Biological Research, POB 19, Nes Zionna

KUPFER, Ephraim
Head, Library, Israel Defence Forces, MPO 1002

KURT, Mrs. Alice
Nuclear Research Center - Negev, POB 9001, Beer Sheva

KUTTEN, Ahron
Librarian, Central Library, The Technion, Haifa

LASSMAN, Mrs. Marilyn
Document Acquisitions Department, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

*LEBOWITZ, Abraham
Director, Elyachar Library, The Technion, Haifa

LEVANON, Itschak
Head of Research Authority, Tel Aviv University, Ramat Aviv

LIPSHITZ, Mrs. Yehudit
Acquisitions Division, The Library, University of the Negev, POB 2053 Beer Sheva

MADGAR, Mrs. Naomi
Tel Aviv University, Ramat Aviv

MAGEEN, Mrs. Binah
Librarian, University of the Negev, POB 2053, Beer Sheva

MAISEL, Mrs. Chana
Armament Development Agency, Ministry of Defence, POB 962, Haifa

ISRAEL, con'd.

MAITLIS, Elina
Head, Division of Information and Documentation, Israel Defence Forces, MPO 2145/A

MANTSUR, Ovadia
Head of Planning Division, National Insurance Foundation, 13 Weizmann Avenue, Jerusalem

MARGOLIN, Mrs. Gerda
Head, Library and Export Documentation Center, Ministry of Commerce and Industry, 9 Carlibach Street, Tel Aviv

MAYERSON, Mrs. Hava
Head of Library, Faculty of Civil Engineering, The Technion, Haifa

MEDWIED, Mordecai
Head, Department of Documentation, Israel Mining Industries, Institute for Research and Development, POB 313, Haifa

MEISNER, Mrs. Erna
Division of Classification, The Library, Soreq Nuclear Research Center, Yavne

MELAMED, Maj. Rachamin
Central Library, Israel Air Force, MPO 2704

MENDRZYCKI, David
Librarian, Faculty of Engineering, The Technion, Haifa

MISCH, Mrs. Miriam
Librarian, Israel Institute for Biological Research, POB 19, Nes Zionna

MOELLER, Mrs. Lies
Head, Acquisitions, The Library, University of Haifa, POB 5354, Haifa

MULDABI, Mrs. Claudia
Head, Reader Services Division, Central Library, Tel Aviv University, Ramat Aviv

NEWMAN, Mrs. Malka
Librarian, University of the Negev, POB 2053, Beer Sheva

OFER, Kurt D.
Armament Development Agency, Ministry of Defence, POB 962, Haifa

ISRAEL, con'd.

*OFFENBACHER, Erwin
Department of Reprography, Hebrew University, Jerusalem

PALDI, Mrs. Ruth
Chief Librarian, Fish Culture Research Station, POB 699, Haifa

PARDO, Mrs. Anita
Librarian, Department of Meteorology Library, Hebrew University, Jerusalem

PELED, Mrs. Ilana
Life Sciences Library, Tel Aviv University, 155 Herzl Street, Tel Aviv

PLATO, Rafael
Chief Librarian, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

PORAT, Shlomo
Administrative Advisor, 4 Zamenhof Street, Herzelia B

PORATH, Mrs. Erna
Life Science Library, Tel Aviv University, 155 Herzl Street, Tel Aviv

PRICE, Ronald
Editor, Desalination Abstracts, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

PUMPIAN, Uzi
Bank of Israel, POB 780, Jerusalem

RABINOVITZ, David
Center for Interdisciplinary Technological Forecasting, Tel Aviv University, Ramat Aviv

RAVIV, Miss Tili
Head, Library, Holon Institute of Technology, POB 305, Holon

RINOTT, Mrs. Rachel
Building Research Station, The Technion, Haifa

ROMM, Mrs. Myriam
Librarian, Mathematics Department Library, Bar Ilan University, Ramat Gan

ISRAEL, con'd.

ROSENHECK, Mrs. Alma
Chief Librarian, Wix Library, Weizmann Institute of Science, POB 26, Rehovot

ROSENTAL, Mrs. Leah
Librarian, Faculty of Agriculture, POB 12, Rehovot

ROTHSCHILD, Mrs. Hadassah
Head of Library, Institute for Training of Social Workers, Jerusalem

ROTHSCHILD, Dr. Jacob
Director, Library School, Hebrew University, Jerusalem

ROZEI-ZUCKER, Mrs. Lola
Librarian, Mifal Mivtsa Co. Ltd. POB 10, Gadera

SCHULMAN, Alexander
Librarian, Scientific Liaison Office, Prime Minister's Office, POB 20125, Tel Aviv

SEGAL, Amihud
Information Expert, Armament Development Agency, Ministry of Defence, POB 962, Haifa

SEVER, Shmuel
Head, Library, University of Haifa, POB 5345, Haifa

SHACHAR, Mrs. Sarah
Documentation and Information Center, Office of the Chief Scientist, Ministry of Defence, Tel Aviv

SHALEV, Mrs. Rachel
Chief Librarian, Tadiran Electronic Industries, POB 648, Tel Aviv

SHALEV, Yitshak
Documentation and Information Center, Office of the Chief Scientist, Ministry of Defence, Tel Aviv

SHANAN, Israel
Head, Department of Organization and Methods (O&M), Ministry of Defence, POB 7076, Tel Aviv

ISRAEL, con'd.

SHAVIT, Arie
Information Scientist, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

SHAVIT, Gershon
Nuclear Research Center - Negev, POB 9001, Beer Sheva

SHEIKIN, Mrs. Irit
National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

SHIFTAN, Mrs. Judith
Ass't. Director, Library, Holon Institute of Technology, POB 305, Holon

SHIOVITZ, Y.
Librarian, American-Israeli Paper Mills Ltd., POB 20181, Tel Aviv

SHNEOR, Mrs. Zina
Librarian, Ministry of Housing, Tel Aviv

SIEGELBAUM, Mrs. Edna
Head of Catalog Department, Central Library, The Technion, Haifa

SILBER, Mendel
Israel Aircraft Industry, Lod Airport

SILBERBERG, Arieh
Librarian, Wix Library, Wiezmann Institute of Science, POB 26, Rehovot

SILBERMAN, Mrs. Yael
Assistant to Director, Central Library, The Technion, Haifa

SIMON, Mrs. Gerti
Head of Library, Israel Institute of Productivity, 4 Henrietta Szold Street, Tel Aviv

SOPHER, Mrs. Miriam
Librarian, Israel Institute for Biological Research, POB 19, Nes Zionna

SPIRA, Mrs. Lea
Librarian, Bet Harofe, Haifa

ISRAEL, con'd.

STECKEL, Dr. Fritz
Researcher, Isotope Department, Weizmann Institute of Science, POB 26, Rehovot

STEIKOVSKY, Mrs. Menucha
Librarian, Computer Sciences Library, The Technion, Haifa

STEINBERG, Ignatz
Industrial Research Center, POB 311, Haifa

STEINER, Paul
Aeronautical Engineer, Information Scientist, Israel Aircraft Industries, Lod Airport

STIMAN, Joel
Programmer, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

STRITMAN, Harry R.
Field Director, Library of Congress Office, American Embassy, Tel Aviv

TAL, Naftali
Ministry of Commerce and Industry, 9 Carlibach St., Tel Aviv

TAL-NIR, Mrs. Sara
Document Acquisitions Department, National Center of Scientific and Technological Information, Prime Minister's Office POB 20125, Tel Aviv

TELEM, Moshe
Tel Aviv University, Ramat Aviv

TI, Alfred
Makhteshim, POB 60, Beer Sheva

TURA, Peretz
Faculty of Agriculture, POB 12, Rehovot

VAN DEN BERG, Mrs. Yael
Head, Division of Information for Industry, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

ISRAEL, con'd.

*VILENTCHUK, Mrs. Lydia
Head, Division of Training and Publications, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

VOLOBRINSKY-KATZ, Mrs. Galia
86 Haroeh Street, Ramat Gan

VOLSHONUK, Joseph
Documentation and Information Center, Office of the Chief Scientist, Ministry of Defence, Tel Aviv

WEIL, Mrs. Susane
Head, Technical Information Division, Soreq Nuclear Research Center, Yavne

WEINER, Haim
Chemicals and Phosphates Ltd., POB 1428, Haifa

WEIS, Mrs. Eva
Head, Forestry Division Library, Volcani Institute of Agricultural Research, POB 6, Beit Dagan

WEISS, Mrs. Magda
Librarian, Building Center of Israel, Haifa

WIDRO, Mrs. Clara
Ministry of Communication, POB 29107, Tel Aviv

WIENER, Miss Pnina
Librarian, Armament Development Agency, Ministry of Defence, POB 962, Haifa

*WOLLMAN, Peretz
Head, Division of Information Systems Development, National Center of Scientific and Technological Information, Prime Minister's Office, POB 20125, Tel Aviv

WURZBURGER, Daniel P.
Armament Development Agency, Ministry of Defence, POB 962, Haifa

YAFFE, Haim
Head, Acquisitions Division, Central Library, Tel Aviv University, Ramat Aviv

ISRAEL, con'd.

YAKVEL, Eliahu
Systems Analyst, Ministry of Defence, POB 7076, Tel Aviv

YATZIV, Shmuel
Director, Haifa Port, POB 539, Haifa

YEHUDAI, Mrs. Diana
Head, Catalog Dept., Central Library, Tel Aviv University, Ramat Aviv

YISRAELI, Reuven
Librarian, Armament Development Agency, Ministry of Defence, POB 962, Haifa

YUDKISS, Dr. Hedda
The Israel Science Teaching Center, Hebrew University, Jerusalem

ZADOKS, Mrs. Annette
Central Library, The Technion, Haifa

ZAITEK, G.
Elta Electronic Industries Ltd., POB 330, Ashdod

ZEIRA, Meir
Ministry of Defence, POB 7055, The Kirya, Tel Aviv

ZINGEREVITZ, Moshe D.
Head, Data Processing Department, Tel Aviv University, Ramat Aviv

ZUCKERMANN, Avraham
Manager, Central Library and Information Service, Israel Aircraft Industries, Lod Airport

ZUPNIK, Mrs. Malka
Head of Library, Mekorot Water Co. Ltd., 9 Lincoln Street, Tel Aviv

ITALY

FANGMEYER, Hermann
Joint Research Center, CCR Euratom, ISPRA (Va)

LUXEMBOURG

*BUNTROCK, Dr. Herbert
Center for Information and Documentation (CID), E.E.C., 29 rue Aldringer

MEXICO

MAGALONI de BUSTAMANTE, Mrs. Ana
Head, Internal Documentation Service, National Council for Science and Technology, POB 20-033, Mexico City

NETHERLANDS

HASPERS, Dr. J.H.
Library, Technological University Delft, POB 98, Delft

LOOSJES, Prof. Th. P.
Librarian, Library of the State Agricultural University, Generaal Foulkesweg 1a, Wageningen

VAN LOEN, Dr. A.
Managing Scientific Director, NOBIN, Burg. v. Karnebeeklaan 19, The Hague

NORWAY

*DISCH, A.
Director, Norwegian Industries Development Association (SNI), Forskningsveien 1, Oslo 3

ROMANIA

IOANID, Mircea
Editor-in-Chief, Center of Information and Documentation in Social and Political Sciences (CIDSP), 3 Gutenberg St., Bucharest 6

NEGOITA, Constantin V.
National Council of Scientific Research, Roma 32, Bucharest

SOUTH AFRICA

BUTTERLY, Miss Elizabeth M.
Atomic Energy Board, Private Bag 256, Pretoria

KINGWILL, Denys G.
Director of Information and Research Services, South African Council for Scientific and Industrial Research, POB 395, Pretoria

*ZAAIMAN, R.B.
Head, Information Service, South African Iron and Steel Industrial Corporation Ltd., Wagonwheel Circle, POB 450, Pretoria

SWEDEN

*GEZELIUS, Dr. Rolf J.E.
Chief Research Engineer, Research Institute of Swedish National Defence FOA 3, S-104 50 Stockholm 80

*HJERPPE, Roland
Royal Institute of Technology Library, S-100 44, Stockholm 70

SAMUELSON, Dr. Kjell
Royal Institute of Technology, S-100 44, Stockholm 70

SCHWARZ, Dr. Stephen
Head of Research Project, Research Institute of Swedish National Defence, 89 Linnegatan, S-104 50, Stockholm

UHLMANN, Wolfram E.
Senior Scientist, Research Institute of Swedish National Defence, FOA Index, S-104 50, Stockholm 80

SWITZERLAND

JOSEPHY, Miss Brigitte E.
Senior Assistant Librarian, Data Handling Division, CERN-Library, 1211 Geneva 23

MOISSE, E.
Head, Information Department, Battelle Memorial Institute, route de Drize 7, 1227 Carouge, Geneva

UNITED KINGDOM

*AITCHISON, T.M.
Manager, Information Research, INSPEC, Institute of Electrical Engineers (IEE), Savoy Place, London WC2R OBL

AMES, Geoffrey R.
Scientific Secretary, Tropical Products Institute, 56/62 Gray's Inn Road, London WC 1X8LT

BARFOOT, Derek E.
Head of Information Department, N.M. Rothschild & Sons Ltd., New Court, St. Swithin's Lane, London EC4P 4DU

UNITED KINGDOM, con'd.

*BARLOW, D.H.
Director, INSPEC, Institute of Electrical Engineers (IEE), Savoy Place, London WC2R OBL

*BATTEN, Dr. W.E.
Director, United Kingdom Chemical Information Service (UKCIS), Burlington House, London W1V OBN

COOPER, S.P.
Deputy Keeper, N.R.L.S.I. (British Museum), 25 Southampton Buildings, Chancery Lane, London WC2A 1AW

*LIEBESNY, Felix
Director, Transociates Ltd., 179 Chevening Road, London NW6 6DT

ROLFE, Mrs. Elizabeth M.
Information Retrieval Officer, Library, National Maritime Museum, Greenwich, London SE10

*SCHUR, Herbert
Senior Lecturer, Postgraduate School of Librarianship and Information Science, University of Sheffield, Western Bank, Sheffield S10 2TN

*SIMMONS, Phillip
Manager, Technical Development, INSPEC, Institute of Electrical Engineers (IEE), Savoy Place, London WC2R OBL

*VICKERY, Brian C.
Head of Research, ASLIB, 3 Belgrave Square, London SW1

WINESTONE, Leslie
Head of Central Technical Records, CIBA-GEIGY (UK) Ltd., Tenax Road, Trafford Park, Manchester 17

UNITED NATIONS

VINCE, Mrs. Violet
Industrial Information Officer, United Nations Industrial Development Organization (UNIDO), Lerchenfelder-strasse 1, 1070 Vienna, Austria

UNITED NATIONS, con'd.

*WYSOCKI, Adam
Director, Division of Scientific Documentation and Information, UNESCO, Place de Fontenoy, 75 Paris 7^{e}, France

UNITED STATES

ABRAMSON, Seth
New York University Medical Center, 566 First Avenue, New York, N.Y.

*ADKINSON, Dr. Burton W.
Director, American Geographical Society, Broadway at 156th Street, New York, N.Y. 10032

*BORKO, Prof. Harold
Graduate School of Library Service, University of California, Los Angeles, Calif. 90024

FRANK, S.
Manager, Technical Information Section, Scientific Services Department, American Cyanamid Company, 1937 W. Main Street, Stamford, Conn. 06904

*GARFIELD, Dr. Eugene
Director, Institute for Scientific Information, 325 Chestnut Street, Philadelphia, Pa. 19106

*HATTERY, Prof. Lowell H.
School of Government and Public Administration, The American University, Washington, D.C. 20016

*HELLMAN, Prof. Richard
Director, Research Center in Business and Economics, University of Rhode Island, Kingston, R.I. 02881

*HIRSCH, Richard S.
Manager, Behavioral Sciences, Information Sciences Department, IBM Corporation, Research Division, Monterey & Cottle Roads, San Jose, Calif. 95114

UNITED STATES, con'd.

KAUFMANN, Mrs. Shoshana
Gift and Exchange Librarian, Paul Klapper Library, Queens College, City University of New York, Flushing, N.Y. 11367

*KELLER, Mark
Research Specialist in Documentation, Rutgers Center of Alcohol Studies, Rutgers University, New Brunswick, N.J. 08903

*KESSLER, Dr. Myer M.
Director, Technical Information Program (TIP), Massachusetts Institute of Technology, Cambridge, Mass. 02139

*KOCHEN, Prof. Manfred
Professor of Information Sciences, Mental Health Research Institute, University of Michigan, Ann Arbor, Mich. 48104

*KRIEGER, Mrs. Tillie
Research Associate, Association of Research Libraries, National Serials Pilot Project, 1521 New Hampshire Avenue, N.W., Washington, D.C. 20636

LOWRY, W.K.
Director, Libraries and Information Systems, Bell Telephone Laboratories, Murray Hill, N.J. 07974

MEIBOOM, Mrs. Esther R.
Librarian, Martland Hospital Unit, New Jersey College of Medicine and Dentistry, 25 Holly Glen Lane South, Berkeley Heights, N.J. 07922

*PAPIER, Lawrence S.
Program Officer, Research and Program Development, Division of Library Programs, Bureau of Libraries and Educational Technology, U.S. Office of Education, Washington, D.C. 20202

PARK, Miss Margaret K.
Information Scientist, Computer Center, University of Georgia, Athens, Ga. 30601

UNITED STATES, con'd.

*PENLAND, Prof. Patrick R.
School of Library Science,
University of Pittsburgh,
135 No. Bellefield Street,
Pittsburgh, Pa. 15213

*SCHAFER, Emil
Systems Analyst, University of
Southern California, University
Computing Center, 1020 W. Jefferson
Blvd., Los Angeles, Calif. 90007

SCHLEIFER, Prof. Harold B.
Chief, Acquisitions Division,
Herbert H. Lehman College
Library, The City University of
New York, Bedford Park Blvd.
West, Bronx, N.Y. 10468

UNITED STATES, con'd.

*SESSIONS, Prof. Vivian S.
Director, Center for Advancement
of Library-Information Science,
City University of New York,
Graduate Center, 33 W. 42nd
Street, New York, N.Y. 10036

WOLFERMAN, Miss Nancy
Instructor, Department of
Library Science, Mount Sinai
School of Medicine of the City
University of New York, 5th
Avenue at 100th Street, New
York, N.Y. 10029

name index

* Paper

** Oral Presentation

author affiliation index

432